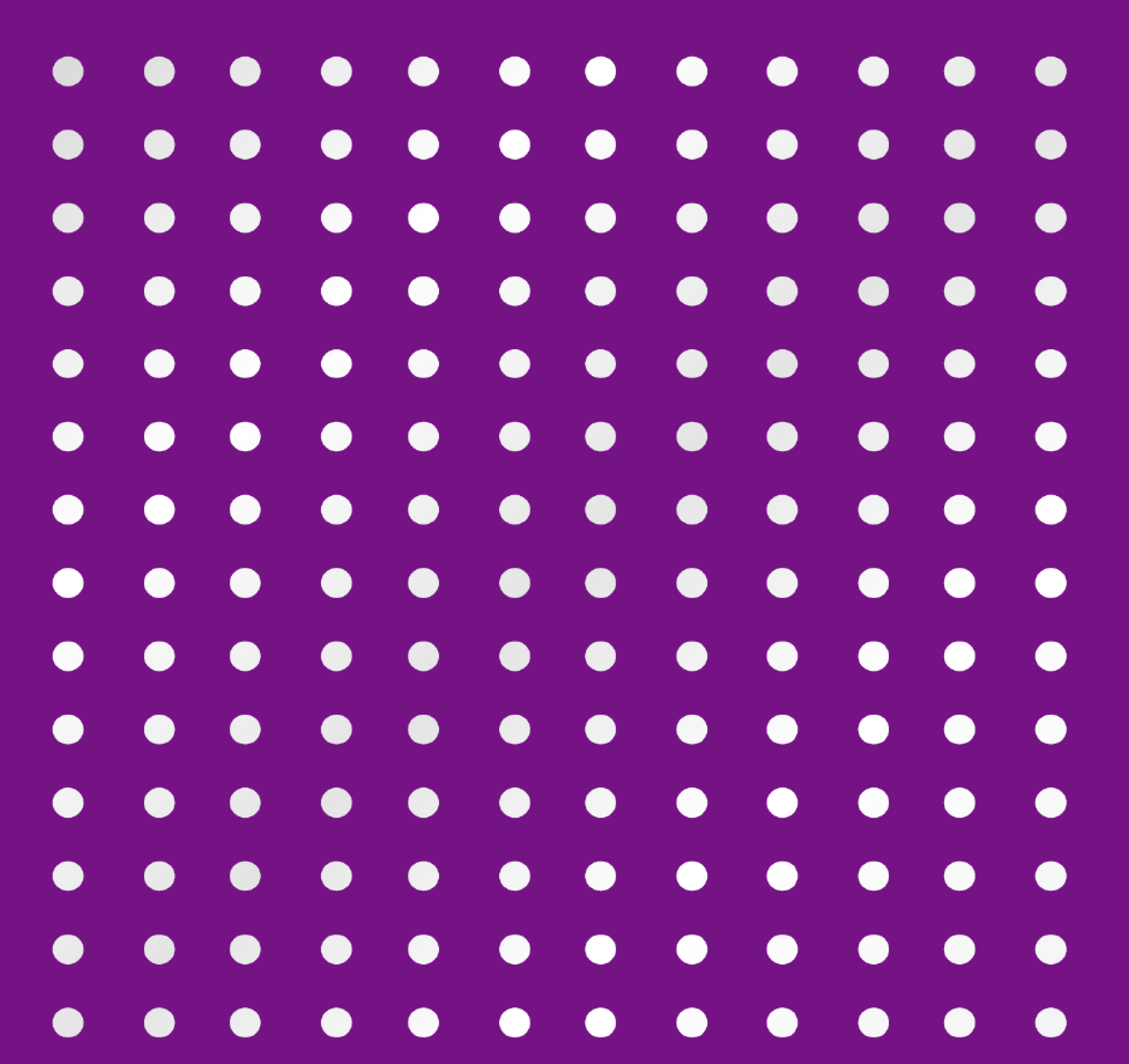
高等院校信息技术规划教材

数字内容安全原理与应用

彭飞 龙敏 刘玉玲 编著
李仁发 主审



清华大学出版社

高等院校信息技术规划教材

数字内容安全原理与应用

彭　飞　龙　敏　刘玉玲　编著
李仁发　主审

清 华 大 学 出 版 社
北　京

内 容 简 介

本书全面介绍了数字内容安全技术的起源、研究发展和应用。全书共分为10章，内容包括绪论、信息加密技术、消息认证与数字签名、信息隐藏与数字水印、数字取证技术、文本内容安全、数字图像内容安全、数字音频内容安全、数字视频内容安全和数据库安全。

本书适合作为信息安全专业本科高年级学生以及研究生的专业课教材，也可供从事信息安全专业技术人员阅读参考。

图书在版编目（CIP）数据

数字内容安全原理与应用/彭飞等编著. --北京：清华大学出版社，2012.7
（高等院校信息技术规划教材）
ISBN 978-7-302-28429-1

Ⅰ. ①数… Ⅱ. ①彭… Ⅲ. ①信息安全－高等学校－教材 Ⅳ. ①TP309

中国版本图书馆CIP数据核字(2012)第060477号

责任编辑：白立军　顾　冰
封面设计：傅瑞学
责任校对：白　蕾
责任印制：王静怡

出版发行：清华大学出版社
网　　址：http://www.tup.com.cn，http://www.wqbook.com
地　　址：北京清华大学学研大厦A座　　**邮　　编：**100084
社 总 机：010-62770175　　**邮　　购：**010-62786544
投稿与读者服务：010-62776969，c-service@tup.tsinghua.edu.cn
质量反馈：010-62772015，zhiliang@tup.tsinghua.edu.cn
印 刷 者：北京世知印务有限公司
装 订 者：三河市新茂装订有限公司
经　　销：全国新华书店
开　　本：185mm×260mm　　**印　　张：**18.25　　**字　　数：**435千字
版　　次：2012年7月第1版　　**印　　次：**2012年7月第1次印刷
印　　数：1～2000
定　　价：39.50元

产品编号：045555-01

前言 Foreword

随着信息技术的发展，数字内容已成为信息的重要表现形式。由于数字内容在互联网上使用的便捷性大大超过了传统模拟形式的信息内容，其应用的广度和深度还在不断增加，数字内容产业已初见规模。然而，数字内容在给人们生活和工作带来便利的同时，也面临着严重的安全威胁。这些威胁主要包括数字内容的非法复制和传播，导致重要信息泄露、数字资产被盗窃；数字内容的非授权篡改，严重影响正常工作进行；数字内容的伪造，导致系统混乱，以至造成各种负面影响；数字内容的可用性，由于非法数据或非正常数据等导致其他数字内容的无法正常和有效使用。安全问题已逐渐成为制约数字内容推广应用的主要瓶颈之一。因此，提高全社会的安全意识和加强信息安全专业知识的教育是保障数字内容产业健康、稳步、快速发展的前提和基础。

数字内容安全是当前信息安全领域的一个重要研究领域，其相关技术还在不断完善。本书的作者在数字内容安全领域开展了一些教学和研究工作，并深感数字内容安全领域的重要性和良好的发展前景。作者结合自己所在单位信息安全专业本科生和相关方向研究生培养的实际情况，编著和出版本书作为专业课程教材。

全书共分为10章，其中第1～5章主要为原理方面的介绍；第6～10章是应用方面的介绍。第1章介绍数字内容的特征、功能以及分类等基本概念，分析数字内容所面临的威胁，并介绍数字内容安全的研究内容与发展历程。第2章介绍密码学的基本原理，主要包括古典密码学、对称密码技术、公钥密码技术以及一些新兴的密码技术(如混沌密码技术与量子密码技术等)，并列出一些经典的密码算法。第3章介绍消息认证与数字签名的基本概念、消息认证的模式与认证方式、单向Hash函数与消息认证码的基本原理、常用的数字签名及一些认证的方法和技术。第4章介绍信息隐藏与数字水印的基本原理，主要包括信息隐藏与数字水印技术的基本概念、空域和变换域的信息隐藏技术、数字水印技术以及信息隐藏与数字水印的发展与应用等。第5章介绍数字取证的基本原理与相关技

术，主要包括数字取证的技术分类、数字内容篡改取证、数字内容来源取证以及数字内容隐秘分析取证，并介绍一些经典的取证案例与取证方法。第6章介绍文本信息的基本概念与文本内容的安全技术，具体包括文本内容加密、文本水印及文本隐写分析技术等。第7章针对数字图像的特点，介绍数字图像以及数字图像内容的相关概念，对数字图像加密技术、数字图像水印技术以及数字图像隐写分析技术进行深入的阐述。第8章对数字音频内容安全的有关概念和方法进行介绍，主要包括数字音频内容加密、数字音频隐写与水印等方面。第9章对数字视频内容安全的有关概念和方法进行介绍，主要包括数字视频内容加密、数字视频隐写与水印、数字视频隐写分析技术与数字视频取证等方面的知识。第10章介绍数据库的基本特性以及数据库所面临的安全威胁，对当前数据库安全技术进行全面的介绍。具体包括数据库的机密性、完整性、访问控制以及安全管理等方面的知识。每章末均给出了适量的思考题作为巩固所学内容之用。

本书作为教材适合于48～64学时的教学，建议的教学方式为课堂讲授与实验相结合，教师可根据书上的练习题，指导学生进行编程或仿真实验，通过对原理和应用算法的实验，进一步加深学生对所学内容的理解。

本书适合作为信息安全专业本科高年级学生以及研究生的专业课教材，也可供从事信息安全专业的技术人员和研究人员阅读参考。

本书作者多年来一直从事信息安全的教学和研究工作，本书也是网络与信息安全湖南省重点实验室全体师生多年从事数字内容安全研究工作成果的结晶。

本书由彭飞负责编写，全书由龙敏和刘玉玲负责整理修改。在本书的编写过程中，陈丽、朱小文、李洪淋、刘娟、李姣婷等研究生参与了部分资料收集与整理工作；湖南大学信息科学与工程学院李仁发教授对本书进行了认真细致的审阅并提供了宝贵的修改意见和建议；清华大学出版社为本书的出版提供了帮助；此外本书的编写还得到了湖南大学信息科学与工程学院赵欢教授的大力支持。在此对他们表示由衷的感谢。

数字内容安全是一门正在发展中的学科，对本书的编著是作者在该领域的一次尝试，由于作者水平有限，书中难免存在疏漏和错误之处，望读者提出宝贵意见，以方便作者日后修改和完善。

作　者

2012年4月

目录 Contents

第1章 chapter 1

绪　论

本章学习目标

随着信息技术的发展，数字内容已成为信息的重要表现形式。由于互联网络的不安全性，数字内容的安全问题开始引起广泛的关注。本章介绍数字内容的特征、功能以及分类等基本概念，分析数字内容所面临的威胁，并介绍了数字内容安全的研究内容与发展历程。

通过本章的学习，应掌握以下内容：

(1) 数字内容的特征、功能以及分类。

(2) 数字内容所面临的威胁及其分类。

(3) 数字内容安全的研究内容与发展的历程。

1.1　数字内容的基本概念

信息是人类社会最重要的资源之一，几乎人类的一切活动都依赖于信息的获取与处理。在现代社会里，信息技术的发展程度已成为衡量一个国家或民族是否进步的重要指标。“信息”一词有着悠久的历史，早在两千多年前的西汉，即有“信”字出现。“信”常可作消息来理解。但对于“信息”一词而言，至今还没有一个公认的定义。从信息的本质来看，它实际上是指事物在相互作用中所“刻画”出的记录。信息的记录方法和社会技术的进步密不可分。古人从“结绳记事”、在龟甲与兽骨上刻画象形文字、在青铜器上铸字、使用木简竹简作为文字载体，到纸张记录，每一次信息记录方法的改变，都是当时社会进步的一个重要标志。进入20世纪中叶以来，随着计算机技术与数字化技术的发展，越来越多的信息开始以数字化的方式存在，为了使敏感的数字化信息内容安全可靠，必须保证数字内容的安全。

1.1.1　数字内容的概念与特征

所谓数字内容，“就是以数字形式存在的文本、图像、声音等信息，它可以存储在如光盘、硬盘等数字载体上，并通过网络等手段传播”。从数字内容的定义来看，它包含如下三个方面的含义：

(1) 数字内容是信息的一种表现形式。也就是说,信息的概念更加广泛,数字内容也隶属于信息的范畴,它只是信息的一种表现形式而已。相对于其他信息的表现形式,数字内容的不同之处在于,数字内容是以数字化的方式存在的。

(2) 数字内容的记录载体是数字化设备。与以往采用麻绳、龟甲与兽骨、青铜器、木简竹简和纸张不同,数字内容记录在数字化设备中,如光盘、U 盘、硬盘以及各种类型的存储卡等。与此同时,存储在数字化设备中的数字内容,通常需要专门的设备才能进行读取。

(3) 数字内容的传播手段是网络。数字内容是可传播的,数字内容只有通过传播才能体现出它的有用性。对于数字内容而言,其传播的手段主要是网络,相对于其他手段,数字内容的传播速度更加快捷。

数字内容是当前信息记录的主要手段,但它自身不能独立存在,它必须依附于某种物质载体。与信息一样,数字内容来源、数字内容归宿以及数字内容的传播信道是组成数字内容的三大要素。数字内容来源是数字内容创建的发源地或出处。数字内容归宿是数字内容的接收者。数字内容的传播信道是数字内容传递的通道,是数字内容来源与数字内容归宿之间的联系纽带。

1.1.2 数字内容的分类

随着数字化技术的发展,数字内容的内涵日益丰富,主要包括数字音像、科学出版、远程教育、动漫游戏、金融信息、政府公告、网络博客、网络论坛、短信彩信、彩铃音乐等,涉及教育、科学、金融、文化、娱乐、商业、通信等多个领域。围绕着这些数字内容的开发制作、传递配送和消费使用,一个影响全社会的大规模的产业链正在形成。

从数字内容的表现形式来看,主要包括数字化的文本、图像、图形、音频、视频等形式。就数字文本而言,比较常见的有电子文档、网络新闻、电子邮件、即时通信、博客、微博等;图像、图形则包含栅格图像(如 JPEG、BMP 等格式的图像)与矢量图形(如 CAD、3ds Max、Coredraw 等图形)。此外,音频、视频也是目前在新闻与娱乐中最为常见的数字内容形式。

从技术方面来讲,数字内容开发、数字内容传递和数字内容安全是组成数字内容的三大支柱。数字内容开发一方面与文化创意和艺术创造紧密结合,同时也与图像、音频、视频、Web 2.0 等技术不可分割;随着宽带技术的发展,数字内容传递正在由传统的离线配送向互联网在线传递和移动传递的方向急剧转变,网络门户、搜索引擎、无线宽带、移动交互等技术成为数字内容传递的核心技术;从一般的信息安全的概念出发,数字内容安全主要应保证内容的隐私性、完整性和真实性。

1.1.3 数字内容的特性

数字内容是一种以电子形式存在的数据,通常是集文本、图像、图形、音频与视频于一体的综合信息,其主要特性如下所示。

(1) 数字化: 在此之前的信息内容几乎都是以模拟的方式进行存储和传播,而数字

内容则是以比特的形式通过数字化设备进行存储、处理和传播的。

(2) 交互性：在模拟领域中，要实现交互性是非常困难的。但在数字内容中，“人机交互作用”则成为可能，故也是数字内容的一个显著特点。

(3) 多样性：主要是指数字内容的表现形式的多样化。人们可以通过视觉、听觉、触觉等多种方式产生、接收数字内容；数字内容通常是技术与艺术的融合，且具有趣味性。

(4) 集成性：主要表现为数字内容通常是多种媒体信息(如文本、图像、音频、视频等)的集成，就是将各种媒体信息按照一定的规则构成一个有机的数字内容整体，用来表现某种信息，使得信息以更为形象的方式进行传播。

(5) 易复制/分发性：人们也可以借助数字技术和互联网，免费并且没有任何质量损失地批量复制和发行数字内容或数字产品。

1.1.4　数字内容相关技术

与数字内容相关的技术范围较广，它是多种学科和多种技术交叉的领域，其主要技术范畴包括以下内容。

(1) 数字内容的表示与操作：包括数字化文字的处理、数字音频处理、数字图像处理、数字视频处理等。

(2) 数字内容压缩：包括通用压缩编码、专用压缩编码技术(声音、图像、视频)等。

(3) 数字内容的存储：包括光盘存储、移动存储、网络硬盘存储等。

(4) 数字内容的管理：包括数字内容管理、数字内容的版权保护等。

(5) 数字内容传输：包括网络传输技术、移动传输技术、流媒体技术、P2P 技术等。

(6) 数字内容的安全：包括保证数字内容的保密性、完整性、可验证性、抗抵赖性、可用性等方面的信息安全技术。

1.2　数字内容面临的威胁与分类

网络技术的飞速发展使得数字内容在互联网上使用的便捷性大大超过了传统的模拟形式的信息内容，数字内容在给人们生活和工作带来便利的同时，也同时面临着严重的安全威胁。

1.2.1　数字内容面临的威胁

数字内容主要包括文档材料、图纸、语音、视频、程序源代码等以电子形式存在的数据，它们所面临的威胁主要包括：

(1) 数字内容的非法复制和传播，导致重要信息泄露、数字资产被盗窃。

(2) 数字内容的非授权篡改，严重影响正常工作进行。

(3) 数字内容的伪造，导致系统混乱，以造成各种负面影响。

(4) 数字内容的可用性，由于非法数据或非正常数据等导致其他数字内容无法正常和有效地使用。

1.2.2 威胁的分类

根据数字内容所面临的威胁,可以将威胁分为主动攻击和被动攻击两类。

主动攻击是指攻击者对数字内容进行某些修改,或者生成一个假的数字内容。它包括非授权的篡改、伪造、内容重放、拒绝服务、伪装等,通常主动攻击较容易被发现。

被动攻击则指攻击者不对数字内容进行任何的改变,只是通过收集通信内容,对其进行分析来获取数字内容中的信息,其攻击方法包括嗅探、信息收集等攻击方法。相对于主动攻击,被动攻击的检测十分困难,但是对这些攻击进行阻止是可能的。

1.3 数字内容安全技术

针对数字内容所面临的安全威胁,数字内容安全技术应运而生。数字内容安全技术是伴随着数字化技术以及网络技术的发展而发展的。

1.3.1 数字内容安全技术的发展历程

数字内容安全技术的发展与信息安全技术的发展是密切相关的,根据不同数字内容安全技术的特征,可分为如下三个阶段。

1. 基于密码术的数字内容安全技术

在这一阶段,数字内容安全主要体现为数字内容的通信安全,通常采用密码技术(如对称密钥密码、公开密钥密码、单向 Hash 函数、数字签名等)保证数字内容的机密性、完整性、可用性和不可否认性。但是,此类方法无法阻止某些被动攻击,如攻击者可以进行通信流量的分析,得到通信的双方以及通信内容的长度。

2. 基于信息隐藏与数字水印的数字内容安全技术

针对基于密码术的数字内容安全技术的不足,研究人员提出了基于信息隐藏与数字水印的数字内容安全技术。该类技术通常通过将重要信息(如版权信息、机密信息等)嵌入到没有安全要求的载体中,通过隐藏重要信息的存在性确保了信息的安全。通过基于信息隐藏与数字水印的数字内容安全技术,可保证载体的版权,内容的完整性。但由于要在载体上加载额外的信息,此类方法通常都会给载体带来一定程度的失真,影响载体的视听效果,严重时甚至会影响到载体的可用性。

3. 基于数字取证的数字内容安全技术

针对基于信息隐藏与数字水印的数字内容技术的不足,研究人员提出了基于数字取证的数字内容安全技术。该类技术通过分析载体的特性(如统计特性、物理特性、环境特性等)来判断载体的真实性或来源。此类技术不需要在载体中加入额外信息,是当前数字内容安全技术中的一个重要研究内容。

上述三类数字内容安全技术的侧重点各有不同,却均有各自的特色。在实际应用中,任何一类技术均无法解决数字内容的所有安全问题,需要三类技术协作实现。

1.3.2 数字内容安全的研究内容

数字内容安全研究的内容主要包括数字内容加密/解密、数字内容信息隐藏、数字内容取证等。

1. 数字内容加密/解密

数字内容加密就是按确定的加密变换方法(加密算法)对需要保护的数字内容(也称为明文)作处理,使其变换成为难以识读的数据(密文)。其逆过程,即将密文按对应的解密变换方法(解密算法)恢复出现明文的过程称为解密。

为了使加密算法能被许多人共用,在加密过程中又引入了一个可变量,即加密密钥。这样,不改变加密算法,只要按照需要改变密钥,也能将相同的明文加密成不同的密文。

加密的基本功能包括:防止不速之客查看机密的数据文件,防止机密数据泄露或被篡改;防止特权用户(如系统管理员)查看私人数据文件,使入侵者不能轻易地查找一个系统的文件等。

2. 数字内容信息隐藏

信息隐藏是将秘密消息隐藏在其他消息中,这样,真正存在的秘密就被隐藏了。通常发送者将秘密信息隐藏在大家耳熟能详的信息载体中,如人民日报的社论、Internet 上广为流传的图片、流行音乐或电影等。

信息隐藏是继加密技术之后,保护数字内容的又一强有力的工具。信息隐藏与传统的信息加密的明显区别在于,传统的加密技术以隐藏信息的内容为目的,使加密后的文件变得难以理解,而信息隐藏是以隐藏秘密信息的存在性为目标。所以科学技术的发展使信息隐藏技术在信息时代成为新的研究热点。它既发扬了传统隐藏技术的优势,又具有了现代的独有特性。

3. 数字内容取证

功能强大的多媒体编辑软件使得数字图像和音视频数据等数字内容的处理变得简单,尽管多数人对数字内容的修改只是为了增强表现效果,但也存在有人出于各种目的传播经过精心伪造的数字图像和音视频数据。篡改和伪造的数字图像和音视频一旦被用于媒体报道、科学发现、保险和法庭证物等,将会对政治、军事和社会的各方面产生严重的影响。因此,需要一种客观、公正、能够澄清事实真相的验证技术,数字内容取证正是为这一目的而提出的。

数字内容取证通常按以下两个原理工作:

(1) 通过对数字内容特征进行分析来判断多媒体内容的完整性、原始性和真实性。

(2) 通过对残留在数字内容内部的设备印迹以及数字信号处理后的噪声进行分析来追溯数字内容数据的来源。

根据应用场合不同，目前国内外数字内容取证研究主要围绕以下五个方面展开：

① 数字内容的篡改检测；

② 数字内容的来源辨识；

③ 多媒体设备的成分取证；

④ 数字内容数据的真实性鉴定；

⑤ 数字内容取证的可靠性。

就媒体类型方面而言，数字图像仍是目前数字取证技术的主要研究对象。

思　考　题

1.1　数字内容有哪些表现形式？各具有什么特点？

1.2　简述主动攻击与被动攻击的特点，并以一种数字内容形式为例列举主动攻击与被动攻击现象。

1.3　简述数字内容目前所面临的安全威胁。

1.4　简述当前数字内容安全的主要研究内容。

参考文献

[1]　刘清堂，陈迪．数字媒体技术导论．北京：清华大学出版社，2008．

[2]　王育民，张彤，黄继武．信息隐藏-理论与技术．北京：清华大学出版社，2006．

[3]　周琳娜，王东明．数字图像取证技术．北京：北京邮电大学出版社，2008．

[4]　胡向东，魏琴芳．应用密码学．北京：电子工业出版社，2009．

[5]　王丽娜，郭迟，李鹏．信息隐藏技术实验教程．武汉：武汉大学出版社，2004．

[6]　冯登国．密码学原理与实践．北京：电子工业出版社，2007．

第 2 章 chapter 2

信息加密技术

本章学习目标

密码技术是保护数字内容安全的一个重要手段。本章将介绍密码学的基本原理，主要包括古典密码学、对称密码技术、公钥密码技术以及一些新兴的密码技术(如混沌密码技术与量子密码技术等)，并列出了一些经典的密码算法。

通过本章的学习，应掌握以下内容：

(1) 古典密码技术的基本原理与分类。

(2) 对称密码技术与 DES、AES 算法。

(3) 对称密码分析技术。

(4) 公钥密码技术与 RSA、ElGamal、ECC。

(5) 混沌密码技术与量子密码技术的发展现状。

2.1 密码学基础

信息加密技术是利用数学或物理手段，对电子信息在传输过程中和存储体内进行保护，以防泄露的技术。保密通信、计算机密钥、防复制软盘等都属于信息加密技术。它是对付各种安全威胁最强有力的工具。

本章将介绍一些密码学中的基础知识和常见的密码学技术。首先介绍了几种常见的古代加密技术及加密算法的使用环境；其次讨论了对称加密，对称加密是公钥密码产生之前唯一的一种加密技术，主要用来提供机密性服务，目前仍有着十分广泛的应用背景；接下来的公钥密码技术，其非对称的独立密钥使得其在消息的保密性、密钥分配和认证领域有着重要的意义；最后简单介绍了近年来新兴的密码技术。本章的学习为后面几章的内容打下基础。

一个密码或者密码体制用于加密数据，原始数据称为明文(plaintext)，通过加密(encryption)对明文进行编码形成密文(ciphertext)，下面再通过解密(decryption)将密文恢复成明文，在密码体制中加密和解密要用到的密钥(key)分别是加密密钥和解密密钥。研究各种加密方案的学科称为密码编码学，研究破译密码获得消息的学科称为密码分析学。传统密码体制模型如图 2-1 所示。

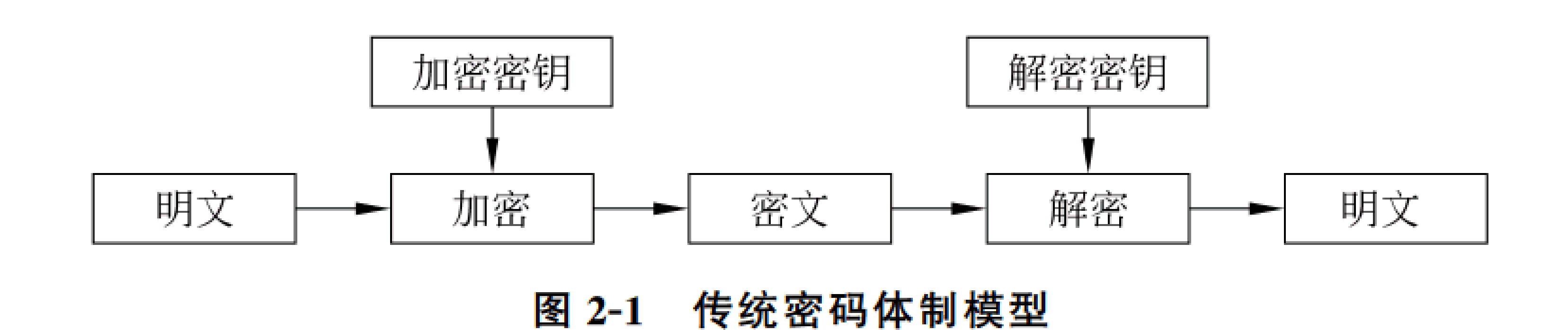


图 2-1 传统密码体制模型

2.2 古典密码技术

在计算机出现前，密码学由基于字符的密码算法构成。不同的密码算法是字符之间互相代换或者是互相之间换位，好的密码算法结合了这两种方法，每次进行多次运算。现在事情变得复杂多了，但是原理还是没有发生变化。不同之处在于算法是对比特而不是对字母进行变换，实际上这只是字母表长度上的改变，从 26 个元素变为 2 个元素。大多数好的密码算法仍然是代替和换位的组合。

本节将介绍 4 种古典密码，每一种都有其独特的地方。

(1) 最古老的密码体制。

(2) 代替密码。

(3) 置换密码。其中也包含现代密码学中一些重要思想，我们还将讨论经典的电码本译码，因为很多现代密码学都可视为等价的“电子”密码本。

(4) 讨论唯一的被证明是安全的密码体制——一次一密密码。

2.2.1 代替密码

代替密码是古典密码中常用到的两种基本处理技巧之一，它在现代密码学中依然得到了广泛的应用。所谓代替，就是将明文中的字母用其他字母、数字或符号所取代的一种方法。常见的代替密码包括单表代替密码、多表代替密码和一次一密。

1. 单表代替密码

单表代替密码对明文中的所有字母都使用同一个映射，即 $\forall p \in P, f: P \to C, c = f(p)$。为了确保解密的正确性，通常要求映射 f 是一一映射的。提到单表代替密码就不得不先说一下凯撒(Caesar)密码。凯撒密码作为一种最为古老的对称加密体制，在古罗马的时候都已经很流行，其基本思想是：通过把字母移动一定的位数来实现加密和解密。例如，如果密钥是把明文字母的位数向后移动三位，由此可见，位数就是凯撒密码加密和解密的密钥。

表 2-1 给出的仅为向后移动三位的凯撒移位，但显然从 1～26 个位置的移位我们都可以使用，将凯撒密码通用化，可以得到如下移位代替密码。

1) 移位代替密码

设：$P=C=K=Z_{26}$，这里，P、C、K、Z_{26} 分别表示明文空间、密文空间、密钥空间和 26 个整数(对应的 26 个英文字母)组成的空间。对于任意大小 $k \in K$，可以得到加密过程

表 2-1　凯撒密码明文密文对照表

明文	a	b	C	d	e	f	g	h	i	j	k	l	m
对应数字	0	1	2	3	4	5	6	7	8	9	10	11	12
密文	D	E	F	G	H	I	J	K	L	M	N	O	P
明文	n	o	p	q	r	s	t	u	v	w	x	y	z
对应数字	13	14	15	16	17	18	19	20	21	22	23	24	25
密文	Q	R	S	T	U	V	W	X	Y	Z	A	B	C

如下：

$$E_k(p) = p + k(\mathrm{mod}\ 26) = c \in C \qquad (2\text{-}1)$$

其中，p 为明文，c 为密文，k 为密钥。

解密过程如下：

$$D_k(c) = c - k(\mathrm{mod}\ 26) = p \in P \qquad (2\text{-}2)$$

例 2-1　当 $k=3$ 时，即为凯撒密码，如表 2-1 所示。

则若明文为：p=casear cipher is a shift substitution 时，密文为：

c=FEVHDU FLSKHU LV D VKLIW VXEVWL WXWLRQ

解密时只需要用密钥 $k=3$ 的加密密钥对密文 c 进行解密运算就可以恢复出原文。

这种密码是将明文字母表中字母位置下标与密钥 k 进行模 26 加法运算的结果作为密文字母位置下标，相应的字母即为密文字母。

2）乘法代替密码

已知：$p=c=k=z_{26}$，k 是满足 $0<k<n$ 的正整数，要求 k 与 n 互素。

加密算法如下：

$$c = E(k,p) = (pk)(\mathrm{mod}\ n) \qquad (2\text{-}3)$$

解密算法如下：

$$p = D(k,c) = k^{-1}c(\mathrm{mod}\ n) \qquad (2\text{-}4)$$

注意：要求 k 与 n 互素原因是仅当 $\gcd(k,n)=1$ 时，才存在两个整数 x 和 y 使得 $xk+yn=1$，才有 $xk\equiv 1 \bmod n$，进而有 $p\equiv xc \bmod n$，明文和密文才是一一对应的，密码才能正确解密。

例 2-2　英文字母表 $n=26$，$k=9$。则有乘法代替密码的明文字母对应表，如表 2-2 所示。

表 2-2　乘法密码明文密文对照表

明文	a	b	c	d	e	f	g	h	i	j	k	l	m
密文	A	J	S	B	K	T	C	L	U	D	M	V	E
明文	n	o	p	q	r	s	t	u	v	w	x	y	z
密文	N	W	F	O	X	G	P	Y	H	Q	Z	I	R

对照上表，若明文为：p=multiplicative cipher，其对应的密文为：c=EYVPUFVU-SAPUHK SUFLKX。

3）仿射密码

乘法密码和加法密码相结合便构成仿射密码。仿射密码是一个线性变换。对于 $p=c=k=z_{26}$，且 $K=\{(a,b)\in z_{26}\times z_{26},\gcd(a,26)=1\}$，对于任意的 $k=(k_1,k_2)\in K$，加密算法如下：

$$c=E(k,p)=k_1p+k_2(\bmod 26) \tag{2-5}$$

解密算法如下：

$$p=D(k,c)=k_1^{-1}(c-k_2)(\bmod 26) \tag{2-6}$$

式中的−1 表示“逆”。显然，当 $k_1=1$ 时，仿射密码对应为凯撒密码。仿射密码共有 26×12=312 个可能的密钥，其中 12 是满足 $\gcd(a,26)=1$ 的 a 的个数。

例 2-3 设 $k=(k_1,k_2)=(5,3)$，可以计算得到 $5^{-1}(\bmod 26)=21$，仿射密码的加密函数为：$c=5p+3(\bmod 26)$；相应的解密函数为：$p=21(c-3)(\bmod 26)=21c-11(\bmod 26)$。

若要加密明文 Cipher，首先转换字母 C，i，p，h，e，r 成数字 2，8，15，7，4，17，然后进行加密：

$$5\times\begin{pmatrix}2\\8\\15\\7\\4\\17\end{pmatrix}+\begin{pmatrix}3\\3\\3\\3\\3\\3\end{pmatrix}=\begin{pmatrix}13\\43\\78\\38\\23\\88\end{pmatrix}\bmod 26=\begin{pmatrix}13\\17\\0\\12\\23\\10\end{pmatrix}=\begin{pmatrix}\mathrm{N}\\\mathrm{R}\\\mathrm{A}\\\mathrm{M}\\\mathrm{X}\\\mathrm{K}\end{pmatrix}$$

即在该密钥下，Cipher 经仿射加密后得到的密文是 NRAMXK。

解密：

$$21\times\begin{pmatrix}13\\17\\0\\12\\23\\10\end{pmatrix}-\begin{pmatrix}11\\11\\11\\11\\11\\11\end{pmatrix}=\begin{pmatrix}262\\346\\-11\\241\\472\\199\end{pmatrix}\bmod 26=\begin{pmatrix}2\\8\\15\\7\\4\\17\end{pmatrix}=\begin{pmatrix}\mathrm{C}\\\mathrm{I}\\\mathrm{P}\\\mathrm{H}\\\mathrm{E}\\\mathrm{R}\end{pmatrix}$$

可见，原始消息 Cipher 已得到恢复。

单表代替密码通常其密钥空间很小，无法抵抗穷举搜索攻击。此外，它没有将明文字母出现的统计概率掩盖起来，容易遭受频率分析攻击。这里所说的频率分析攻击是指在某种语言中，由于不同字符出现频率的差异所呈现出来的统计规律。

2. 多表代替密码

隐藏字母出现的频率分布并提高代替密码强度的一种方法是采用多个密文字母表，使密文中的每一个字母有多种可能的字母来代替，多表代替密码有多个单字母密钥，每一个密钥被用来加密一个明文字母。第一个密钥加密明文的第一个字母，第二个密钥加

密明文的第二个字母。在所有的密钥用完后,密钥又再循环使用。

已知明文序列为 $p=p_1p_2\cdots$,$f=f_1f_2\cdots$为映射序列,则对应的密文为:

$$C=E(k,p)=f_1(p_1)f_2(p_2)\cdots \tag{2-7}$$

若 f 是非周期的无限序列,则相应的密码称为非周期多表代换密码。这类密码,对每个明文字母都采用不同的代换表(或密钥)进行加密,称作一次一密密码(one-time pad cipher),这是一种理论上唯一不可破的密码,这种密码对于明文的特点可实现完全隐蔽,但由于需要的密钥量和明文信息的长度相同而限制其广泛使用。

在多表代换下,原来明文中的统计特性通过多个表的平均作用而被隐蔽了起来。多表代换密码的破译要比单表代替密码的破译难得多。

但是多表代换中的平均结果会使密文的统计特性与明文的统计特性明显不同,随着多表代换周期的加大,这种差别也就更加明显,从此入手就可以破译多表代换密码。

Vigenère 密码、Playfair 密码、滚动密钥密码、弗纳姆密码以及 Hill 密码都是这一类密码。

1) Vigenère 密码

Vigenère 密码是最著名的多表代换密码,是由法国密码学家 Blaise de Vigenere 于 1568 年提出的一种密码,它是一种以移位代换为基础的周期代换密码、多表简单加法密码。

Vigenère 密码使用一个词组作为密钥,每一个密钥字母都对应一个代替表。第一个密钥字母用来加密第一个明文字母,第二个密钥字母加密第二个明文字母,等所有密钥字母都使用完后,密钥又再循环使用。

已知明文 $p=p_1p_2\cdots p_n$,m 为一个固定的正整数,对于一个密钥 $k=k_1k_2\cdots k_m$,则加密算法如下:

$$\begin{aligned}C&=E(p,k)\\&=(p_1+k_1(\mathrm{mod}\ 26),p_2+k_2(\mathrm{mod}\ 26),\cdots,p_i+k_i(\mathrm{mod}\ 26),\cdots)\end{aligned} \tag{2-8}$$

解密算法如下:

$$\begin{aligned}P&=D(c,k)\\&=(c_1-k_1(\mathrm{mod}\ 26),c_2-k_2(\mathrm{mod}\ 26),\cdots,c_i-k_i(\mathrm{mod}\ 26),\cdots)\end{aligned} \tag{2-9}$$

Vigenère 密码使用 26 个密文字母表,像加法密码一样,他们是一次将明文字母表循环右移 0,1,2,…,25 位的结果。选一个词组或者短语作为密钥,以密钥字母控制使用哪一个密文字母表。

例 2-4 已知明文 p=polyalphabetic cipher,密钥 k=RADIO,即周期 $d=5$,则

- 明文:p=polyalphabetic cipher;
- 密钥:k=RADIORADIORADI ORADIO;
- 密文:c=GOOGOCPKIPVTLK QZPKMF;

其中,同一明文字母 P 在不同的位置被加密成不同的字母 G 和 P。

2) Playfair 加密算法

Playfair 密码将明文中的双字母组合作为一个单元进行处理,并将每一个单元转换成双字母的密文组合。Plairfair 密码基于一个 5×5 矩阵,该矩阵采用一个关键词作为密钥来构造。构造的方法为:按从左至右,从上至下的顺序依次首先填入关键词中非重复的

字母，然而再将字母表中剩余的字母按顺序填入矩阵(其中字母 I 和 J 被看作是一个字母)。

对于每一对明文 p_1、p_2，其加密方法如下：

① p_1 和 p_2 在同一行时，则密文 c_1 和 c_2 分别是紧靠 p_1、p_2 右端的字母。其中第一列看作是最后一列的右方。

② 若 p_1 和 p_2 在同一列时，则密文 c_1 和 c_2 分别是紧靠 p_1、p_2 下方的字母。其中第一行看作是最后一行的上方。

③ 若 p_1 和 p_2 不在同一行，也不在同一列时，则密文 c_1 和 c_2 是由 p_1 和 p_2 确定的矩形的其他两角的字母，并且 c_1 和 p_1、c_2 和 p_2 同行。

④ 若 $p_1 = p_2$，则插入一个字符(如 Q)于重复字母之间。

⑤ 若明文字母为奇数时，将空字母 Q 加在明文的末端。

例 2-5 密钥是：EXAMPLE FOR PLAYFAIR，则构造的字母矩阵如表 2-3 所示。

表 2-3 字母矩阵表

E	X	A	M	P
L	F	O	R	Y
I/J	B	C	D	G
H	K	N	Q	S
T	U	V	W	Z

如果明文是 p = chinese student，先将明文每两个分为一组：

ch in es es tu de nt

按照加密规则，对应的密文为：

IN CH PH PH UV IM HV

Playfair 密码相对于单表替换密码有很大的进步，主要体现在两个方面：

(1) 由于是双字母组合，共有 26×26=676 种组合的可能，识别双字母组合要更为困难。

(2) 各个字母组合的频率比单字母呈现出大得多的范围，导致频率分析的难度加大。即便如此，Playfair 密码还是相对容易攻破，因为在密文中仍然存在许多明文语言的结构可被密码分析者利用。

3) 滚动密钥密码

对于周期多表代替密码，保密性将随周期 d 的加大而增加，当 d 的长度和明文一样长时就变成了滚动密钥密码。如果其中所采用的密钥不重复就是一次一密体制。一般密钥可取一本书或一篇报告作为密钥源，可由书名，章节号及标题来限定密钥的起始位置。

4) 弗纳姆密码

当字母表字母数 $q=2$ 时，滚动密钥密码就变成了弗纳姆密码。

选择随机二元数字序列作为密钥，以 $k=k_1k_2\cdots k_i\cdots(k_i\in F_2)$ 表示，明文字母编程二元向量后也可以表示为二元序列 $m=m_1m_2\cdots m_i\cdots(m_i\in F_2)$，则加密过程就是将 k 和 m 的相应位逐位的模 2 相加，即：

$$c_i = m_i \oplus k_i,\quad i = 1,2,\cdots \tag{2-10}$$

译码时，用同样的密钥对密文逐位的模 2 加，便可恢复明文的二元数字序列，即：

$$m_i = c_i \oplus k_i,\quad i = 1,2,\cdots \tag{2-11}$$

这种加密方式若使用电子器件实现就是一种序列密码。

5) Hill 密码

Hill 加密算法的基本思想是将 m 个明文字母通过线性变换将它们转换为 m 个密文

字母。解密只要做一次逆变换就可以了。密钥就是变换矩阵本身。假设 $m=3$，则

$$\begin{cases} c_1 = k_{11}p_1 + k_{12}p_2 + k_{13}p_3 \\ c_2 = k_{21}p_1 + k_{22}p_2 + k_{23}p_3 \\ c_3 = k_{31}p_1 + k_{32}p_2 + k_{33}p_3 \end{cases} \tag{2-12}$$

可用列向量和矩阵来表示：

$$\begin{bmatrix} c_1 \\ c_2 \\ c_3 \end{bmatrix} = \begin{bmatrix} k_{11} & k_{12} & k_{13} \\ k_{21} & k_{22} & k_{23} \\ k_{31} & k_{32} & k_{33} \end{bmatrix} \begin{bmatrix} p_1 \\ p_2 \\ p_3 \end{bmatrix} \tag{2-13}$$

即加密过程为：

$$\boldsymbol{C} = \boldsymbol{KP} \bmod 26 \tag{2-14}$$

其中，$\boldsymbol{C}$ 和 $\boldsymbol{P}$ 代表密文和明文向量，$\boldsymbol{K}$ 是密钥矩阵。

解密则为：

$$\boldsymbol{P} = \boldsymbol{K}^{-1}\boldsymbol{C} \tag{2-15}$$

例 2-6 加密明文为 july，密钥矩阵为：$\boldsymbol{k}=\begin{bmatrix} 11 & 3 \\ 8 & 7 \end{bmatrix}$，则加密是先将明文分为两个组 ju(9,20)和 ly(11,24)。加密算法如下：

$$\boldsymbol{c}_1 = \begin{bmatrix} 11 & 3 \\ 8 & 7 \end{bmatrix}\begin{bmatrix} 9 \\ 20 \end{bmatrix} = \begin{bmatrix} 3 \\ 4 \end{bmatrix} \quad \boldsymbol{c}_2 = \begin{bmatrix} 11 & 3 \\ 8 & 7 \end{bmatrix}\begin{bmatrix} 11 \\ 24 \end{bmatrix} = \begin{bmatrix} 11 \\ 22 \end{bmatrix}$$

因此，加密后的密文为：DELW。

解密算法如下(密钥矩阵的逆矩阵)：

$$\boldsymbol{k}^{-1} = \begin{bmatrix} 7 & 23 \\ 18 & 11 \end{bmatrix}$$

$$\boldsymbol{p}_1 = \begin{bmatrix} 7 & 23 \\ 18 & 11 \end{bmatrix}\begin{bmatrix} 3 \\ 4 \end{bmatrix} = \begin{bmatrix} 9 \\ 20 \end{bmatrix}, \quad \boldsymbol{p}_2 = \begin{bmatrix} 7 & 23 \\ 18 & 11 \end{bmatrix}\begin{bmatrix} 11 \\ 22 \end{bmatrix} = \begin{bmatrix} 11 \\ 24 \end{bmatrix}$$

因此，解密后可得到原始密文 july。

6) 一次一密

一次一密密码是一种较为理想的加密方案，由 Major Joseph Mauborgne 和 AT&T 公司的 Gilbert Vernam 于 1917 年发明。一次一密乱码本是一个大的不重复的真随机密钥字母集，这个密钥字母集被写在几张纸上，并一起粘成一个乱码本。发方用乱码本中的每一密钥字母准确地加密一个明文字符。加密是明文字符和一次一密乱码本密钥字符的模 26 加法。

每个密钥仅对一个消息使用一次。发方对所发的消息加密，然后销毁乱码本中用过的一页或用过的磁带部分。收方有一个同样的乱码本，并依次使用乱码本上的每个密钥去解密密文的每个字符。收方在解密消息后销毁乱码本中用过的一页或用过的磁带部分。新的消息则用乱码本的新的密钥加密。

如果偷窃听者不能得到用来加密消息的一次一密乱码本，这个方案是完全保密的。给出的密文消息相当于同样长度的任何可能的明文消息。随机密钥序列异或一非随机的明文消息产生一完全随机的密文消息。再强大的计算能力也无能为力。

密钥字母必须是随机产生的。对这种方案的攻击将是针对用来产生密钥序列的那种方法。使用伪随机数发生器是不值得考虑的，它们通常具有非随机性。如果采用真随机源，它就是安全的。

另一个重要的事情是密钥序列不能重复使用。一次一密乱码本的想法很容易推广到二进制数据的加密，只需由二进制数字组成的一次一密乱码本代替由字母组成的一次一密乱码，用异或代替一次一密乱码本的明文字符加法就可以了。为了解密，用同样的一次一密乱码本对密文异或，其他保持不变，保密性也很完善。

但一次一密乱码本存在几个问题。因为密钥比特必须是随机的，并且绝不能重复使用，密钥序列的长度要等于消息的长度。即使解决了密钥的分配和存储问题，还需确信发方和收方是完全同步的。如果收方有一比特的偏移(或者一些比特在传送过程中丢失了)，消息就变成乱的了。另一方面，如果某些比特在传送中被改变了(没有增减任何比特，更像由于随机噪声引起的)，那些改变了的比特就不能正确地解密。再者，一次一密乱码本不提供鉴别。

一次一密乱码本在今天仍有应用场合，主要用于高度机密的低带宽信道。

2.2.2 置换密码

把明文中的字母重新排列，字母本身不变，但其位置改变了，这样编成的密码称为置换密码。最简单的置换密码是把明文中的字母顺序倒过来，然后截成固定长度的字母组作为密文。

例 2-7 明晨 5 点发动反攻。

明文：

MING CHEN WU DIAN FA DONG FAN GONG

密文：

GNOGN AFGNO DAFNA IDUWN EHCN IM

这种技巧对密码分析者来说实在微不足道。一种更复杂的方案是把消息一行一行地写成矩形块，然后按列读出，但是把列的次序打乱，列的次序就是算法密钥。

例 2-8 密钥：4 3 2 1 5 6 7

明文：

a	t	t	a	c	k	P
o	s	t	p	o	n	e
d	u	n	t	i	l	t
w	o	a	m	x	y	z

密文：

TTNAAPTMTSUOAODWCOIXKNLYPETZ

单纯的置换密码因为有着与原文相同的字母频率而被识破，如同列变换所示，密码分析可以直接将密文排列成矩阵入手，再来处理列的位置。双字母音节和三字母音节可

以派上用场。

多步置换密码相对来说安全得多。这种复杂的置换是不容易构造出来的。因此，如果前面的那条消息用相同算法再加密一次，则密文为：

NSCYAUOPTTWLTMDNAOIEPAXTTOKZ

2.3 对称密钥密码技术

在很长一段时间内，密码技术主要应用于军事以及外交等领域，直到1977年，美国国家标准局公布实施了美国数据加密标准(Data Encryption Standard，DES)，军事部门垄断密码的局面被打破，民间力量开始全面介入密码学的研究和应用中。市场上涌现出大量的民用加密产品，常用的加密算法有DES、IDEA、AES等。

2.3.1 基本概念

对称密钥加密又叫专用密钥加密，即发送和接收数据的双方必须使用相同的密钥对明文进行加密和解密运算。对称密钥加密算法主要包括DES、3DES、IDEA、FEAL、BLOWFISH等。对称密钥是双方使用相同的密钥，在网络条件下就要求使用一个安全的信道进行密钥的共享与传递。

对称加密的基本要求包括：

(1) 需要强大的加密算法。算法至少应该满足：即使分析人员知道了算法并能访问一些或者更多的密文，也不能破译出密文或得出密钥。通常，这个要求以更强硬的形式表达出来，那就是即使分析人员拥有一些密文和生成密文的明文，也不能译出密文或者发现密钥，即加密算法应足以抵抗已知明文类型的破译。

(2) 发送方和接收方必须用安全的方式来获得密钥的副本，保证密钥的安全。如果有人发现了密钥，并知道了算法，则使用此密钥的所有通信便都是可读取的。

对称密钥密码技术有两种不同的实现方式，分别是流密码技术与分组密码技术。

2.3.2 流密码技术

流密码的基本思想是利用密钥 k 产生一个密钥流 $k_0k_1k_2\cdots$，并使用如下规则对明文串 $p=p_0p_1p_2\cdots$ 加密：$c=c_0c_1c_2\cdots=E_{k0}(p_0)E_{k1}(p_1)E_{k2}(p_2)\cdots$。密钥流由密钥流发生器 f 产生：$z_i=f(k,\sigma_i)$，这里 σ_i 是加密器中的记忆元件(存储器)在时刻 i 的状态，f 是由密钥 k 和 σ_i 产生的函数。

流密码的滚动密钥 $z_0=f(k,\sigma_0)$ 由函数 f、密钥 k 和指定的初态 σ_0 完全确定。此后，由于输入加密器的明文可能影响加密器中内部记忆元件的存储状态，因此 $\sigma_i(i>0)$ 可能依赖于 $k,\sigma_0,x_0,x_1,\cdots,x_{i-1}$ 等参数。

根据加密器中记忆元件的存储状态 σ_i 是否依赖于输入的明文字符，流密码可进一步分成同步和自同步两种。σ_i 独立于明文字符的叫做同步流密码，否则叫做自同步流密码。由于自同步流密码的密钥流的产生与明文有关，因而较难从理论上进行分析。目前

大多数研究成果都是关于同步流密码的。在同步流密码中，由于 $Z_i=f(k,\sigma_i)$ 与明文字符无关，因而此时密文字符 $y_i=E_{Z_i}(x_i)$ 也不依赖于此前的明文字符。因此，可将同步流密码的加密器分成密钥流产生器和加密变换器两个部分。如果与上述加密变换对应的解密变换为 $x_i=D_{Z_i}(y_i)$，则可给出同步流密码体制的模型如图 2-2 所示。

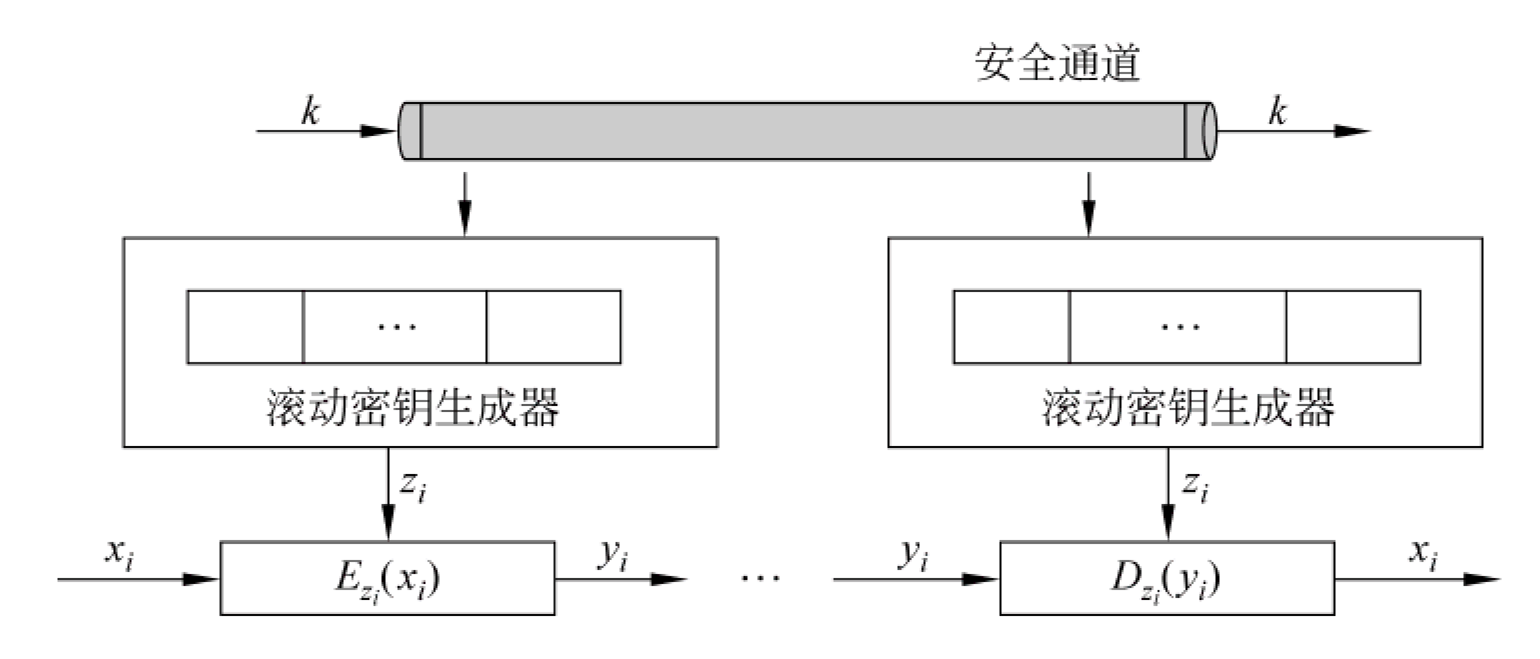


图 2-2　同步流密码体制的模型

实际使用的数字保密通信系统一般都是二元系统，因而在有限域 GF(2)上讨论的二元加法流密码是目前最为常用的流密码体制，其加密变换可表示为 $y_i=z_i\oplus x_i$。实际使用中，密码设计者的最大愿望是设计出一个滚动密钥生成器，使得密钥经其扩展成的密钥流序列具有如下性质：极大的周期、良好的统计特性、抗线性分析、抗统计分析。

下面将详细介绍两种流密码算法：A5/1 和 RC4。这两种算法在当今被广泛应用。A5/1 在 GSM 移动通信中使用，A5/1 算法是基于硬件实现的流密码的代表。RC4 算法在安全套接字 SSL 协议等许多地方有广泛的使用。RC4 是一种特殊的流密码，其软件实现效率非常高。

1. A5/1

A5/1 算法主要应用在 GSM 移动通信中用于保护数据。该算法可以通过代数描述，也可任意使用简单的流程图来描述。这里同时给出这两种描述。

A5/1 使用 X、Y、Z 三个线性移位寄存器 LFSR。寄存器 X 包括 19 位，编号为(x_0，x_1，…，x_{18})。寄存器 Y 包括 22 位，编号为(y_0，y_1，…，y_{21})。寄存器 Z 包括 23 位，编号为(z_0，z_1，…，z_{22})。三个 LFSR 总共包括 64 比特。

密钥 K 同样是 64 位，用于初始化三个寄存器。用密钥填充三个寄存器后，就完成了密码流生成前的准备。在描述密码流之前，首先介绍三个寄存器的详细结构。

对于寄存器 X，每步进行如下操作：

$$\begin{aligned} t &= x_{13} \oplus x_{16} \oplus x_{17} \oplus x_{18} \\ x_i &= x_{i-1}, \quad i = 18,17,16,\cdots,1 \\ x_0 &= t \end{aligned} \tag{2-16}$$

类似地，对于寄存器 Y 和 Z，每步分别进行如下操作：

$$\begin{aligned} t &= y_{20} \oplus y_{21} \\ y_i &= y_{i-1}, \quad i = 21,20,19,\cdots,1 \end{aligned}$$

$$y_0 = t \tag{2-17}$$

和

$$\begin{aligned} t &= z_7 \oplus z_{20} \oplus z_{21} \oplus z_{22} \\ z_i &= z_{i-1}, \quad i = 22,21,20,\cdots,1 \\ z_0 &= t \end{aligned} \tag{2-18}$$

给定三个比特 x、y、z，定义 maj(x,y,z)为“多数投票”函数：即如果 x、y、z 中的多数为 0，则函数返回 0，否则返回 1。

A5/1 使用硬件实现的，每个时钟周期作如下计算：

$$m = \text{maj}(x_8, y_{10}, z_{10}) \tag{2-19}$$

于是寄存器 X、Y、Z 依照如下规则进行处理：

(1) 如果 $x_8 = m$，那么就进行 X 操作。

(2) 如果 $y_{10} = m$，那么就进行 Y 操作。

(3) 如果 $z_{10} = m$，那么就进行 Z 操作。

最后，密钥流比特 s 按照如下关系产生：

$$s = x_{18} \oplus y_{21} \oplus z_{22} \tag{2-20}$$

为了生成一个比特的密钥流的过程看似复杂，但是 A5/1 的硬件实现非常简单，比特产生的速度与时钟速度相当。并且从一个 64 位的密钥可以产生无穷多的密钥流，尽管最终密钥流将出现循环。A5/1 算法可以使用简单的“电码”表示，如图 2-3 所示。

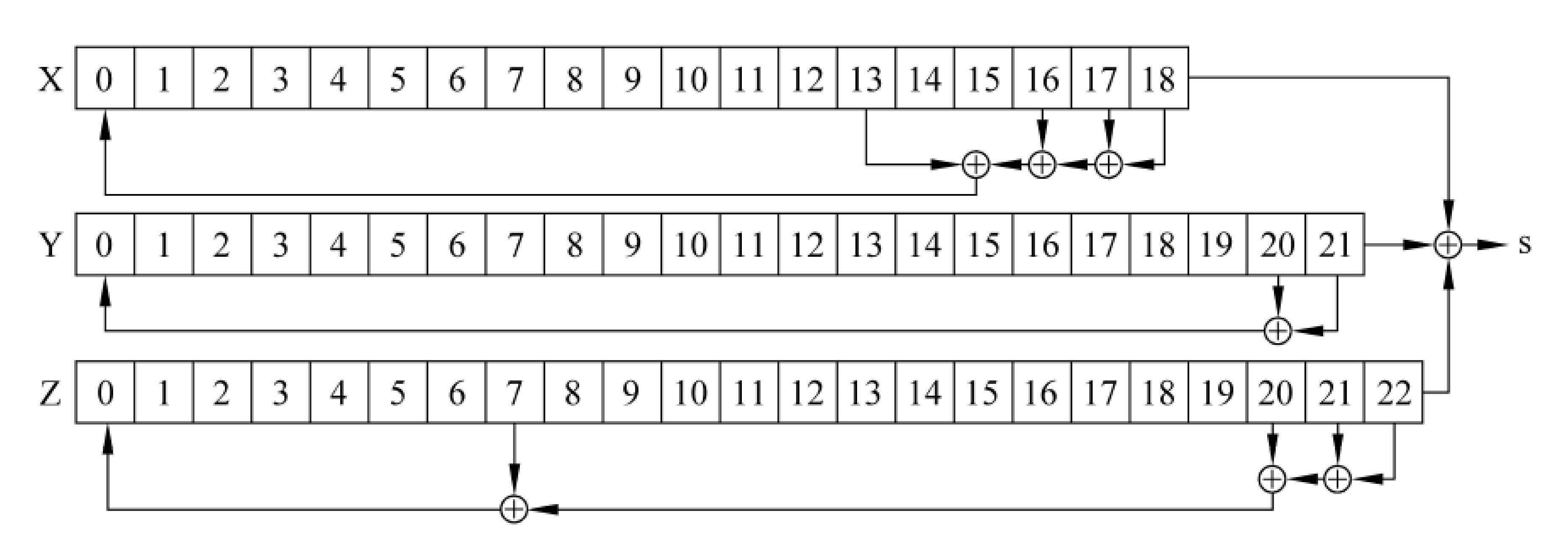


图 2-3 A5/1 密钥流生成

2. RC4

RC4 也是一种流密码，但是它与 A5/1 有很大的不同。RC4 算法专门为软件实现优化，而 A5/1 则是根据硬件实现设计；RC4 每步产生一个密钥字节，而 A5/1 每步仅产生一个密钥流比特。

RC4 算法非常简单，因为从本质上来讲它就是一个包含 256 字节的置换查表在产生密钥流的每一个字节时，所查的表就进行一次修改，表始终都包含{0,1,2,…,255}的置换。

整个 RC4 算法都是基于字节的。算法的第一阶段是对于查表使用的密钥进行初始化，用 key[i]表示密钥，这里 $i=1,2,\cdots,N-1$，每个 key[i]是一个字节，标记为$s[i]$，这里每个 $s[i]$也是一个字节。RC4 的一个特点是，密钥长度可以是 0～256 字节。密钥只在

初始化置换 S 中使用。

置换 S 的初始化过程的伪码如下：

```
for i=0 to 255
    s[i]=i
    k[i]=key[i mod N]
Next i
j=0
for i=0 to 255
    j=(j+s[i]+k[i]) mod 256
    swap(s[i],s[j])
next i
i=j=0
```

初始化阶段完成后，通过下列代码中的算法产生每个密钥流字节。可以用 keystreamByte 表示输出，在加密时与明文做 XOR 运算，解密时与密文做 XOR 运算。RC4 算法的输出同时也可为需要“密码学”伪随机数的应用作为伪随机数生成器使用。

RC4 密钥流字节如下：

```
i=(i+1) mod 256
j=(j+s[i]) mod 256
swap(s[i],s[j])
t=(s[i]+s[j]) mod 256
keystreamByte=s[t]
```

RC4 算法可以被视为自修改的查找表，它非常简单，并且软件实现效率很高。然而对于 RC4 存在可行的攻击方法，但是只要在使用时丢弃生成前 256 字节密钥流，该攻击就不可行。这可以通过在初始化过程中额外添加 256 步来完成，每一步产生 RC4 密码流字节中被丢弃的密钥流字节。

RC4 可以在包括 SSL 在内的很多应用中使用。然而该算法比较过时，没有针对 32 位处理器进行优化。

2.3.3 分组密码技术

分组密码是对称密码的典型代表。即数据在密钥的作用下一组一组地被处理，并且明文和密文的长度通常是相等的，一次对一个明文分组(如 DES 为 64 位)进行加密，而且每次的加密密钥都相同，分组加密的一般结构如图 2-4 所示。

当密钥给定时，对于每一个明文分组，都有唯一的一个密文分组与之对应。因此可以想象有一个非常大的电码本，对每一个可能的明文分组，在电码本中都有唯一与之对应的密文分组。对于大于分组长度的报文，需将其分为若干特定长度的分组，最后一个分组可能需要填充。解密过程也是一次对一个密文分组进行解密。而且每次解密都使用同一个密钥。

分组密码用于短数据(如加密密钥)加密时效果非常理想，但如果同一明文分组在消

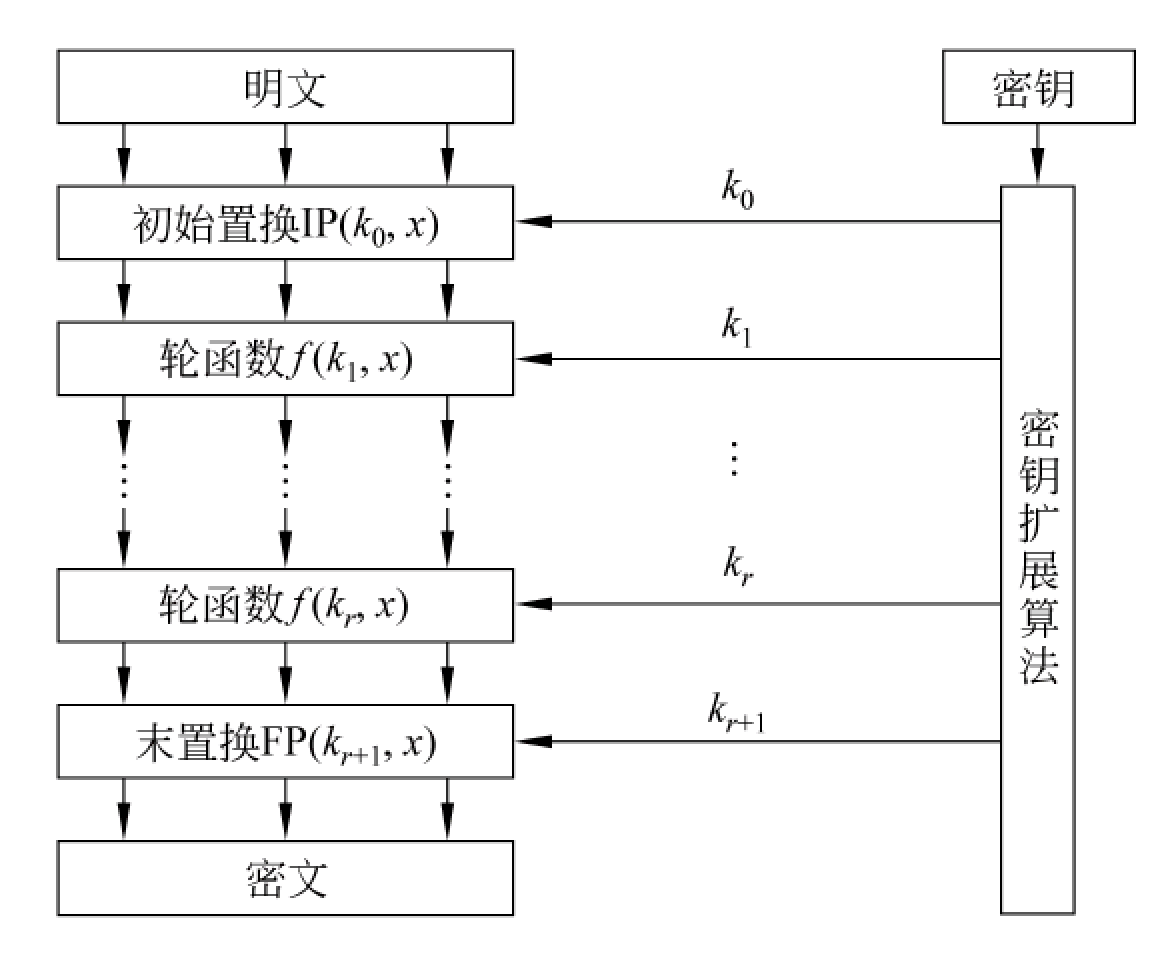


图 2-4 分组加密的一般结构

息中反复出现，产生的密文分组就会相同，不仅容易被攻击者抓住规律猜测攻击，而且在时间上也大大重复了相同的工作。因此，用于长消息加密时可能不够安全。

如图 2-5 所示，给定加密消息的长度是随机的，按特定长度(如 64 位)分组时，最后一组消息长度可能不足 64 位。可以填充一些数字，通常用最后 1 字节作为填充指示符(PI)。它所表示的十进制数字就是填充占有的字节数。数据尾部、填充字符和填充指示符一起作为一组进行加密。

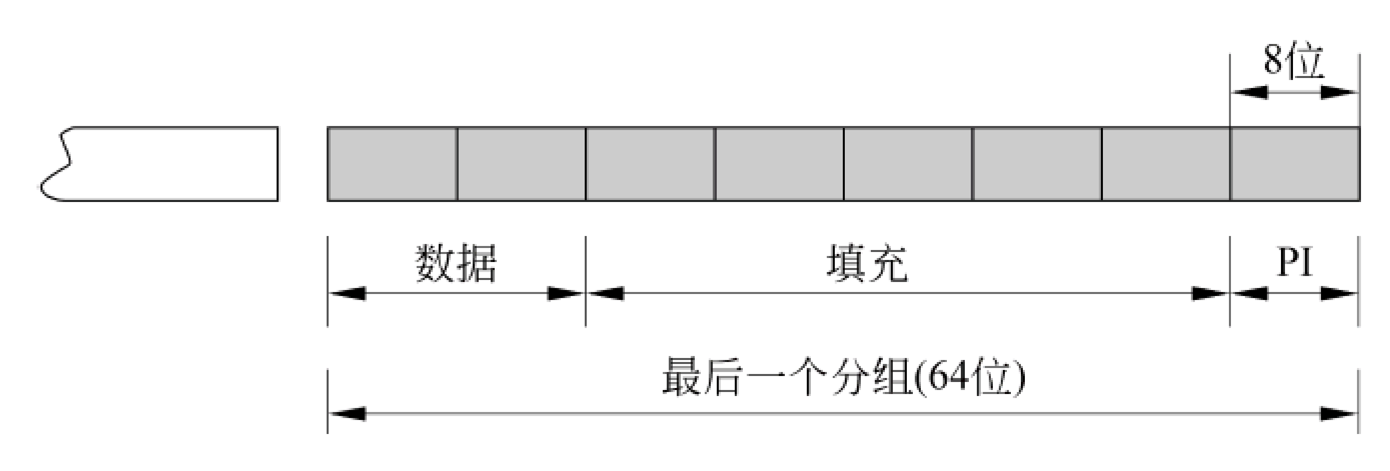


图 2-5 分组密码的消息填充

1. 数据加密算法标准

1973 年，美国国家标准局(National Bureau of Standards，NBS)开始征集一种标准的数据加密标准算法(DES)，以用于非机密性政府机构、商业部门和民间的对非机密的敏感数据进行加密。IBM 公司在 1971 年完成的 LUCIFER 密码(64 比特分组，128 比特密钥)的基础上，改进后成为建议的 DES。1975 年 3 月 17 日，NBS 公布了这个算法，并说明要以它作为联邦信息处理标准，征求各方意见。1977 年 1 月 15 日，建议被批准为联邦标准——FIPSPUB 46，并设计推出了 DES 芯片。1981 年，ANSI 将 DES 作为标准，即 DEA[ANSI X3. 92]。1983 年，ISO 采用 DES 作为标准，即 DEA-1。DES 是一个优秀的对称分组密码算法，直到 2000 年 10 月 2 日 NIST 宣布 AES 算法前，其一直是业界的标准。

DES 是一种分组乘积密码，包括 16 轮迭代。明密文分组长度为 64 位，密钥总长为 64 位，有效长度 56 位，其中第 8，16，…，64 位共 8 位是奇偶校验位。DES 是一种对称运

算，除子密钥使用顺序逆序外，加密和解密算法相同。DES 是一种面向二进制的密码算法，能够加解密任何形式的计算机数据。

DES 的加密算法流程如图 2-6 所示，主要包括三大步骤：

(1) 初始置换 IP：把输入的 64 位数据块的排列顺序打乱，每位数据按照下面换位规则重新组合。$IP(b_1b_2b_3\cdots b_{64})=b_{58}b_{50}\cdots b_7$ 即将输入的第 58 位换到输出的第 1 位，将输入的第 50 位换到输出的第 2 位……输入的第 7 位换到输出的第 64 位，将变换后的数据平分成各 32 位的左右两部分，左部分记为 L_0，右部分记为 R_0，如表 2-4 所示。

(2) 16 轮的轮变换：首先密钥扩展算法将 64 位的输入密钥(称为主密钥 master key)扩展为加解密各轮所需的轮子密钥(sub key)。DES 共需要 16 个轮子密钥，每个轮子密钥有 48 位。对 R_0 实行在轮子密钥 k_1(轮子密钥由密钥扩展算法产生)控制下的变换 f，结果记为 $f(R_0,k_1)$，再与 L_0 做按位异或运算，其结果记为 R_1，R_0 则直接作为下一轮的 L_1，如此循环 16 轮，得到预输出结果 R_{16}、L_{16}。

$$\begin{cases}L_n=R_{n-1}\\R_n=L_{n-1}\oplus f(R_{n-1},K_n)\end{cases},\quad n=1,2,\cdots,16 \tag{2-21}$$

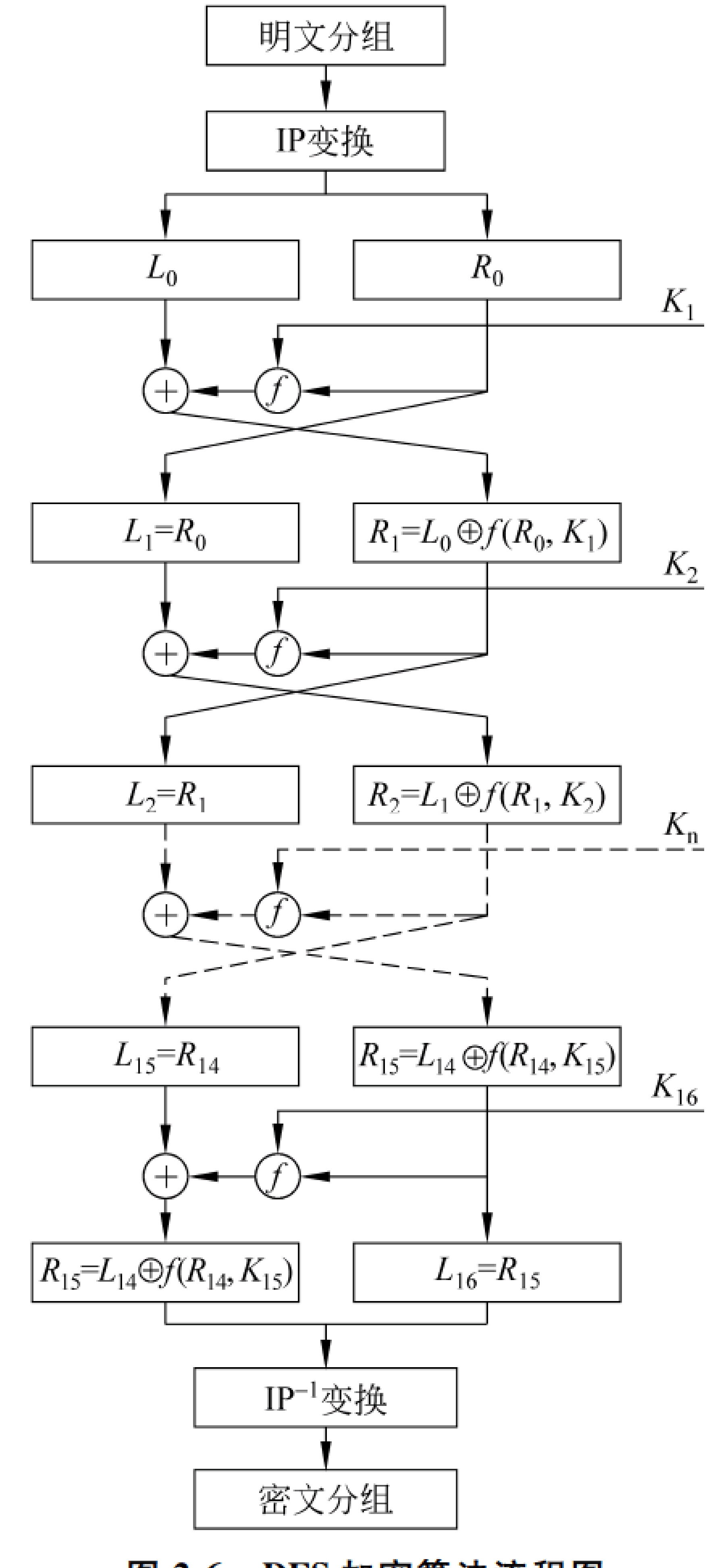


图 2-6 DES 加密算法流程图

f 函数是多个置换函数和替代函数的组合函数，它将 32 位比特的输入变换为 32 位的输出。如表 2-5 所示 32 位的 R 经过扩展变换 E(Expend)后，扩展为 48 位的 $E(R)$，然后与 48 位的轮子密钥 K 进行按位异或。$E(R)=E(b_1b_2b_3\cdots b_{32})=b_{32}b_1\cdots b_1$，输出的第 1 位为输入的第 32 位，输入的第 2 位为输入的第 1 位，输入的第 48 位为输入的第 1 位，如表 2-5 所示。E 的主要作用是增加算法的扩散效果。

表 2-4 IP 置换

58	50	42	34	26	18	10	2
60	52	44	36	28	20	12	4
62	54	46	38	30	22	14	6
64	56	48	40	32	24	16	8
57	49	41	33	25	17	9	1
59	51	43	35	27	19	11	3
61	53	45	37	29	21	13	5
63	55	47	39	31	23	15	7

表 2-5 E 的置换

32	1	2	3	4	5
4	5	6	7	8	9
8	9	10	11	12	13
12	13	14	15	16	17
16	17	18	19	20	21
20	21	22	23	24	25
24	25	26	27	28	29
28	29	30	31	32	1

(3) 逆初始置换 IP^{-1}：逆初始置换 IP^{-1}是初始置换 IP 的逆置换，它将由 L_{16}、R_{16}合并的 64 位数据作为输入，进行换位后得到 64 位的密文输出。$IP^{-1}(b_1b_2b_3\cdots b_{64})=b_{40}b_8\cdots b_{25}$，即将输入的第 40 位换到输出的第 1 位，将输入的第 8 位换到输出的第 2 位……输入的第 25 位换到输出的第 64 位，如表 2-6 所示。

表 2-6 IP^{-1}置换

40	8	48	16	56	24	64	32
39	7	47	15	55	23	63	31
37	6	46	14	54	22	62	30
36	5	45	13	53	21	61	29
35	4	44	12	52	20	60	28
34	3	43	11	51	19	59	27
33	2	42	10	50	18	58	26
32	1	41	9	49	17	57	25

2. 高级加密标准(AES)

从各方面来看，DES 已走到了它生命的尽头。因为其 56 比特密钥实在太小，虽然三重 DES 可以解决密钥长度的问题，但是 DES 的设计主要针对硬件实现，而在当今许多领域，需要用软件方法来实现。在这种情况下，它的效率相对较低。鉴于此，1997 年 4 月 15 日美国国家标准和技术研究所(NIST)发起征集高级加密标准(Advanced Encryption Standard，AES)算法的活动，并成立了 AES 工作组。目的是为了确定一个非保密的、公开披露的全球免费使用的加密算法，用于保护下一世纪政府的敏感信息。也希望能够成为保密和非保密部门公用的加密算法。

AES 是 Rijndael 算法的一个子集，已经由 NIST 通过 FIPS-197 标准化了。AES 算法是 128 位块密码，支持三种不同大小的密钥：128、192 和 256 位。最大优点是可以给出算法的最佳差分特征的概率及最佳线性逼近的偏差的界，由此，可以分析算法抵抗差分密码分析及线性密码分析的能力。

AES 密码算法采用的是代替-置换网络(Substitution-Permutation Network，SPN)结构，每一轮操作由 4 层组成：第 1 层(字节替换)为非线性层，用 S 盒对每一轮中的单个字节分别进行替换；第 2 层(行移位)和第 3 层(列混合)是线性混合层，对当前的状态按行移位、按列混合；第 4 层(密钥加层)用子密钥与当前状态进行字节上的异或，AES 的具体算法结构如图 2-7 所示。

图 2-7(a)给出了算法的整体结构，输入明文 P 与子密钥 K_0 异或，然后经过 R 轮迭代最终生成密文 C，其中第 1 到第 $R-1$ 轮迭代结构如图 2-7(b)所示，第 R 轮与前面各轮稍微有点不同，缺少混合层。

其中，加密轮数与密钥长度的关系如表 2-7 所示。

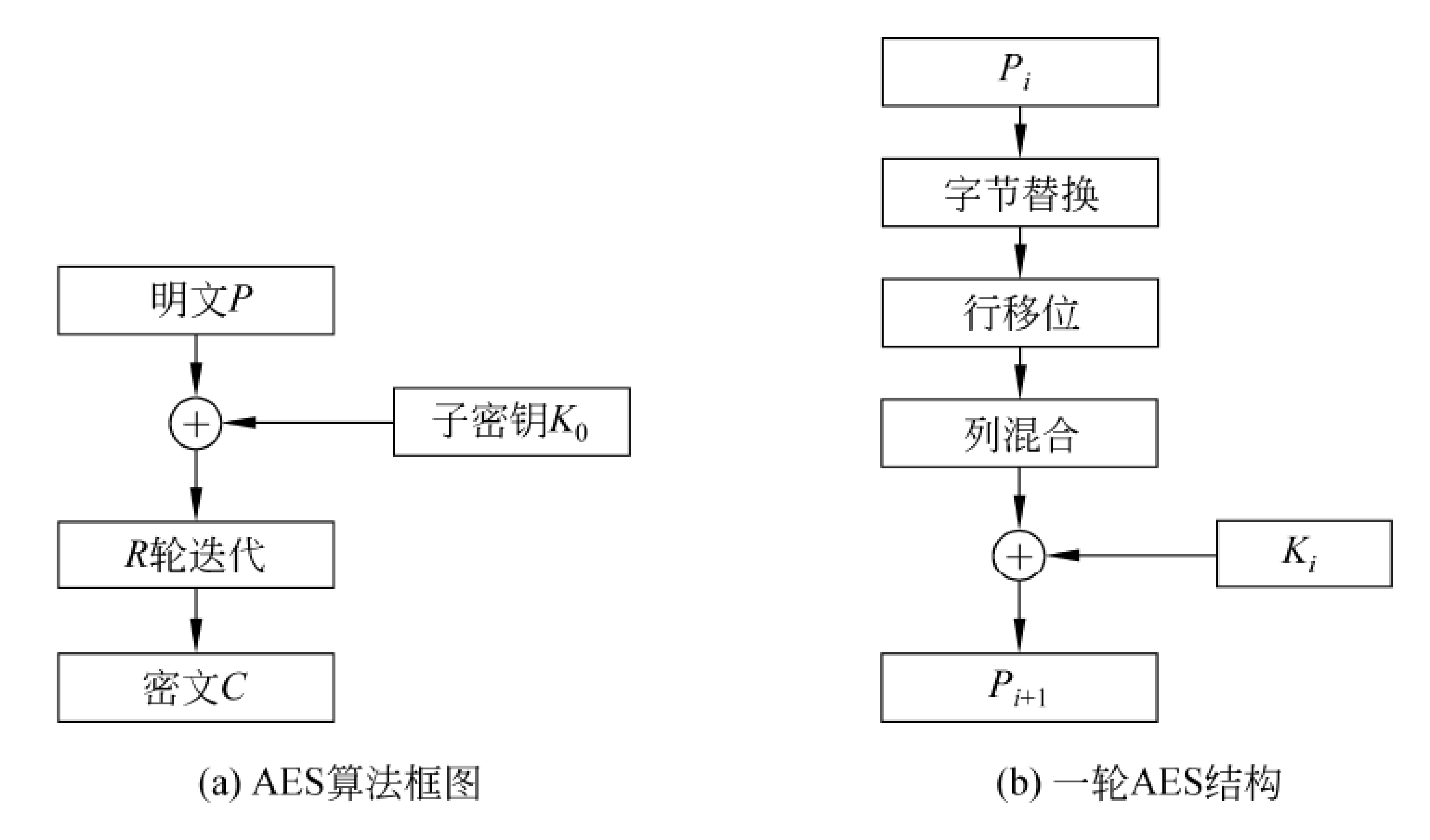


图 2-7 AES 算法结构图

表 2-7 AES 参数

密钥长度/bit	128	192	256
明文分组长度/bit	128	128	128
轮数	10	12	14
每轮密钥长度/bit	128	128	128
扩展密钥长度/B	176	206	240

1）字节替换(SubBytes)

AES 定义了一个 S 盒，State 中每个字节按照如下方式映射为一个新的字节：把该字节的高 4 位作为行值，低 4 位作为列值，然后取出 S 盒中对应行和列的元素作为输出。例如，十六进制数{84}。对应 S 盒的行是 8，列是 4，S 盒中该位置对应的值是{5F}。

S 盒是一个由 16×16 字节组成的矩阵，包含了 8 位值所能表达的 256 种可能的变换。S 盒按照以下方式构造：

① 逐行按照升序排列的字节值初始化 S 盒。第一行是{00}，{01}，{02}，…，{OF}；第二行是{10}，{l1}，…，{1F}等。在行 X 和列 Y 的字节值是{xy}。

② 把 S 盒中的每个字节映射为它在有限域 GF(2^k)中的逆。GF 代表伽罗瓦域，GF(2^8)由一组从 0x00 到 0xff 的 256 个值组成，加上加法和乘法。GF(2^8)$=\dfrac{Z_2[X]}{(x^8+x^4+x^3+x+1)}$。{00}被映射为它自身{00}。

③ 把 S 盒中的每个字节记成(b_8，b_7，b_6，b_5，b_4，b_3，b_2，b_1，b_0)。对 S 盒中每个字节的每位做如下变换：

$$b_i' = b_i \oplus b_{(i+4)\bmod 8} \oplus b_{(i+5)\bmod 8} \oplus b_{(i+6)\bmod 8} \oplus b_{(i+7)\bmod 8} \oplus c_i$$

上式中 c_i 是指值为{63}字节 C 第 i 位，即($c_8c_7c_6c_5c_4c_3c_2c_1c_0$)=(01100011)。符号(′)表示更新后的变量的值。AES 用以下的矩阵方式描述了这个变换：

$$\begin{bmatrix} b_0 \\ b_1 \\ b_2 \\ b_3 \\ b_4 \\ b_5 \\ b_6 \\ b_7 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 & 0 & 1 & 1 & 1 \\ 1 & 1 & 1 & 0 & 0 & 0 & 1 & 1 \\ 1 & 1 & 1 & 1 & 0 & 0 & 0 & 1 \\ 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} b_0 \\ b_1 \\ b_2 \\ b_3 \\ b_4 \\ b_5 \\ b_6 \\ b_7 \end{bmatrix} + \begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \\ 0 \\ 1 \\ 1 \\ 0 \end{bmatrix}$$

2）行位移变换（ShiftRows）

State 的第一行字节保持不变，State 的第二行字节循环左移一个字节，State 的第三行字节循环左移两个字节，State 的第四行循环左移三个字节。变化如图 2-8 所示。

14	0	5d	ab
78	10	C1	fd
31	9	11	3f
28	0b	2a	45

ShiftRows交换 →

14	0	5d	ab
10	C1	fd	78
11	3f	31	9
45	28	0b	2a

图 2-8 ShiftRows 变换

3）列混合变换（MixColumns）

列混合变换是一个替代操作，是 AES 最具技巧性的部分。它只在 AES 的第 0，1，…，$R-1$ 轮中使用，在第 R 轮中不使用该变换。乘积矩阵中的每个元素都是一行和一列对应元素的乘积之和。在 MixColumns 变换中，乘法和加法都是定义在 GF(2^8)上的。State 的每一列($b_{i,j}$)$i=0,\cdots,3$；$j=0,\cdots,L_b$ 被理解为 GF(2^8)上的多项式，该多项式与常数多项式 $a(x)=a_3x^3+a_2x^2+a_1x+a_0$ 相乘并模 $M(x)=x^4+1$ 约化。

这个运算需要做 GF(2^8)上的乘法。但由于所乘的因子是三个固定的元素 02、03、01，所以这些乘法运算仍然是比较简单的（注意，乘法运算所使用的模多项式为 $m(x)=x^8+x^4+x^3+x+1$）。设一个字节为 $b=(b_7b_6b_5b_4b_3b_2b_1b_0)$，则

$b\times$‘01’$=b$；

$b\times$‘02’$=b_6b_5b_4b_3b_2b_1b_0 0$；

$b\times$‘03’$=b\times$‘01’$+b\times$’02’。

注意：加法为取模 2 的加法，即逐比特异或。

写成矩阵形式为：

$$\begin{bmatrix} b_0 \\ b_1 \\ b_2 \\ b_3 \end{bmatrix} = \begin{bmatrix} 02 & 03 & 01 & 01 \\ 01 & 02 & 03 & 01 \\ 01 & 01 & 02 & 03 \\ 03 & 01 & 01 & 02 \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ a_2 \\ a_3 \end{bmatrix}$$

4）轮密钥加变换（Add RoundKey）

轮密钥加变换是 128 位的 State 按位与 128 位的密钥进行 XOR 操作：(b_{0j}，b_{1j}，b_{2j}，

$b_{3j}) \leftarrow (b_{0j}, b_{1j}, b_{2j}, b_{3j}) \oplus (k_{0j}, k_{1j}, k_{2j}, k_{3j})$，对 $j=0,\cdots,R-1$，轮密钥加变换很简单，却影响了 State 中的每一位。密钥扩展的复杂性和 AES 的其他阶段运算的复杂性确保了该算法的安全性。

5）密钥扩展(Key Expansion)

为了防止已有的密码分析攻击，AES 使用了与轮相关的轮常量（Rcon[j]是一个字，这个字的右边三个字节总为 0)防止不同轮中产生的轮密钥的对称性或相似性。AES 在加密和解密算法使用了一个由种子密钥字节数组生成的密钥调度表，AES 规范中称之为密钥扩展。密钥扩展过程从一个原始密钥中生成多重密钥以代替使用单个密钥大大增加了比特位的扩散，在 AES 密钥扩展算法的输入值是 4 字密钥，输出是一个 44 字的一维线性数组。这足以为初始轮密钥扩展过程阶段和算法中的其他 10 轮中的每一轮提供 16 字节的轮密钥。

通过生成器产生 N_r+1 轮密钥，每个轮密钥由 N_b 个字组成，共有 $N_b(N_r+1)$个字 $W[i]$（$i=0,1,\cdots,N_b(N_r+1)-1$）。

在加密过程中，需要 N_r+1 个子密钥，需要构造 $4(N_r+1)$个 32 位字。Rijndael 的密钥扩展方案的伪码描述如下：

```
KeyExpansion(byte key[4* Nk],word w[Nb* (Nr+1)],Nk)
{ //Nk代表以 32位字为单位的密钥的长度,即 Nk=密钥长度/32
begin
  i=0
  while(I<Nk)
  w[i]=word[key[4* i],key[4* i+1],key[4* i+2],key[4* i+3]]
  i=i+1
end while
i=Nk
while(i<Nb* (Nr+1))
   word temp=w[i-1]
   if (I mod Nk=0)
      temp=SubWord(RotWord(temp))xor Rcon[i/Nk]
   else if (Nk=8 and I mod Nk=4)
      temp=SubWord(temp)
end if
w[i]=w[i-Nk] xor temp
i=i+1
end while
end
}
```

其中，key[]和 w[]分别用于存储扩展前，扩展后的密钥。SubWord()、RotWord()分别是与 S 盒的置换和以字节为单位的循环移位。Rcon[i]=(RC[i],'00','00','00')，RC[0]='01'，RC[i]=2 • (RC[i-1])(i>1)。前 10 个轮常数 RC[i]的值（用十六进制表示）如表 2-8 所示，其对应的 Rcon[i]如表 2-9 所示。

表 2-8 前 10 个轮常数 RC[i]的值

i	1	2	3	4	5	6	7	8	9	10
RC[i]	01	02	04	08	10	20	40	80	1b	36

表 2-9 对应的 Rcon[i]的值

I	1	2	3	4	5
Rcon[i]	01000000	02000000	04000000	08000000	10000000
I	6	7	8	9	10
Rcon[i]	20000000	40000000	80000000	1b000000	36000000

输入密钥直接被复制到扩展密钥数组的前四个字中，得到 $w[0]$、$w[1]$、$w[2]$、$w[3]$；然后每次用四个字填充扩展密钥数组余下的部分。在扩展密钥数组中，$w[i]$的值依赖于 $w[i-1]$和 $w[i-4]$($i\geqslant 4$)。

对 w 数组中下标不为 4 的倍数的元素，只是简单地异或，其逻辑关系为：$w[i]=w[i-1]\oplus w[i-4]$(i 不为 4 的倍数)。

对 w 数组中下标为 4 的倍数的元素，采用如下方法计算：

① 将一个字的四个字节循环左移一个字节，即将字$[b_0,b_1,b_2,b_3]$变为$[b_1,b_2,b_3,b_0]$；

② 基于 S 盒对输入字中的每个字节进行 S 代替；

③ 将步骤②的结果再与轮常量 Rcon[i]相异或。

④ 将步骤②的结果与 $w[i-4]$异或。

2.3.4 对称密钥密码的分析方法

密码编码学和密码分析学既对立又统一，正是由于它们的对立性才促进了密码学的发展。密码分析学是在不知道密钥的情况下，恢复出密文中明文信息的方法。根据密码分析者对明文、密文等信息掌握的多少，可以将密码分析分为以下五种情形。

(1) 唯密文攻击：对于这种形式的密码分析，分析者只知道加密算法和待破译的密文。

(2) 已知明文攻击：破译者已知的内容包括加密算法和经密钥加密形成的一个或多个明文-密文对。

(3) 选择明文攻击：破译者除了知道加密算法外，还可以选定明文消息，并可以知道对应的密文。

(4) 选择密文攻击：破译者除了知道加密算法外，还包括自己选定的密文和对应的、已解密的明文。

(5) 选择文本攻击：破译者已知的东西包括加密算法、由密码破译者选择的明文消息和它对应的密文，以及由密码破译者选择的猜测性明文和它对应的已破译的明文。

下面介绍主要的分析方法。

1. 强力攻击法

强力攻击可用于任何分组密码，且攻击的复杂度仅依赖于分组长度和密钥长度。严格地讲，攻击所需的时间复杂度依赖于分组密码的工作效率，其工作效率包括加解密速度、密钥扩展速度、存储空间等。

2. 差分密码分析

差分密码分析是迄今为止已知最有效的攻击迭代密码的方法之一，它利用高概率特征或差分恢复密钥。其基本思想为：通过分析明文对的差值对密文对的差值的影响来恢复某些密钥比特。简单地，随机选取具有固定差分的一对明文，只要它们符合特定的差分条件，甚至可以不必知道它们的值。然后，按照不同的概率，将输出密文中的差分分配给不同的密钥。随着对密文对的分析越来越多，将使最可能的一个密钥显现出来，这样就得到了正确的密钥。差分密码分析最初是针对 DES 加密提出的一种攻击方法，可用于 6 轮以上的 DES 加密。8 轮 DES 需要 2^{14} 个选择明文，10 轮和 14 轮 DES 分别需要 2^{24} 和 2^{39} 个选择明文才能破解。虽然差分密码分析未能破解 16 轮的 DES 加密，但用它破解轮数较低的 DES 还是很成功的。例如，在个人计算机上几分钟就可以破解 8 轮 DES。差分密码分析除了用来攻击 DES 外，也可以被用来攻击其他的密码体制。

3. 线性密码分析

线性密码分析本质上是一种已知明文攻击法，是对 DES 加密方法进行破译的主要方法。这种方法用 2^{21} 个已知明文可以破译 8 轮 DES，用 2^{47} 个明文可以破译 16 轮 DES。在某些情况下，这种方法可用于唯密文攻击。其基本思想是：通过寻找一个给定密码算法的有效的线性近似表达式来破译密码系统。由于每个密码系统均为非线性系统，因此只能寻找线性近似表达式。如果分别将明文的一些位、密文的一些位进行异或运算，然后再将这两个结果进行异或运算，这两个结果的运算结果是一个位，这一位与密钥的一些位进行异或运算的结果相同。这一位就是概率为 P 的线性近似值，在 P 不等于 1/2 前提下，就可以使用该偏差，用得到的明文及相对应的密文便可猜测密钥的位值。得到的明文数据越多，猜测密钥的位置越可靠。概率 P 越大，用同样数据量分析的成功率就越高。

4. 差分-线性密码分析

强力攻击、差分密码分析和线性密码分析是三种对 DES 主要的攻击方法。由于差分密码分析和线性密码分析对于 16 轮的 DES 的分析所需的选择(已知)明文个数太大，所以目前最有效的攻击仍然是强力攻击。而差分—线性密码分析就是对差分密码分析和线性密码分析进行改进，是降低它们复杂度的众多改进之一，它利用的是差分密码分析和线性密码分析相结合的技术。

5. 插值攻击

插值攻击仅对某些密码算法有效，即轮数很少或轮函数的次数很低的算法。如果密

文可以表示成明文的多项式，则插值攻击根据具体条件可以给出等价于加密或解密算法的一个变换，或者恢复出最后一轮的子密钥。该方法利用了拉格朗日插值公式的思想。插值攻击由 Knudsen 和 Jakobsen 提出，如果一个密码算法是固定密钥的低次多项式函数，或项数较少的多项式，其项数可以被估算出来，则通过插值法可以得到其代数表达式，从而恢复出密钥；在改进后的插值攻击中，可以精确地计算出多项式函数的某些项的系数，在利用有限域上傅里叶变换的基础上，也可以求出相应的密钥。另外，如果密文可以作为两个多项式的商，且可以估计出来这两个多项式的项数，那么相应的密钥同样可以恢复出来。插值攻击使用代数函数来代表 S 盒，可以用已知明文攻击法取得此函数的样本点，再用拉格朗日插值法产生。这个代数函数可能是在有限体上的有理函数、多项式函数或二次函数。此函数也可以用选择明文攻击法取得样本点，这样可以简化所使用的代数函数，让攻击效率更高。Thoms Jakobsen 又将机率的概念引入了插值攻击法，通过 MadhuSudan 演算法来改善其对 Reed-Solomon 纠错码的解译能力。如此一来，在明文与密文的内容仅有极少的代数关系时插值攻击也有效。

2.4 公钥加密技术

2.4.1 基本概念

公开密钥算法的思想最早是由当时在美国斯坦福大学的 Diffie 和 Hellman 两人在 1976 年在其论文 *New Direction in Cryptography* 中提出的。但目前最流行的 RSA 算法是 1977 年由 MIT 教授 Ronald L. Rivest，Adi Shamir 和 Leonard M. Adleman 共同开发的，分别取自三名数学家名字的第一个字母来构成的。

1976 年提出的公开密钥密码体制思想不同于传统的对称密钥密码体制，它要求密钥成对出现，一个为加密密钥(e)，另一个为解密密钥(d)，且不可能从其中一个推导出另一个。自 1976 年以来，已经提出了多种公开密钥密码算法，其中许多是不安全的，一些被认为是安全的算法又有许多是不实用的，它们要么是密钥太大，要么密文扩展十分严重。多数密码算法的安全基础是基于数学难题，这些难题专家们认为在短期内不可能得到解决。因为一些问题(如因子分解问题)至今已有数千年的历史了。

公钥加密算法也称非对称密钥算法，用两个密钥：一个公共密钥和一个专用密钥。用户要保障专用密钥的安全；公共密钥则可以发布出去。公共密钥与专用密钥是有紧密关系的，用公共密钥加密的信息只能用专用密钥解密，反之亦然。由于公钥算法不需要联机密钥服务器，密钥分配协议简单，所以极大简化了密钥管理。除加密功能外，公钥系统还可以提供数字签名。非对称密码算法解决了对称密码体制中密钥管理的难题，并提供了对信息发送人的身份进行验证的手段，是现代密码学最重要的发明和进展。

单向和陷门单向函数的概念是公钥密码学的核心，可以说公钥密码体制的设计就是陷门单向函数的设计。

给定任意两个集合 X 和 Y。函数 $f: X \to Y$ 称为单向的，如果对每一个 x 属于 X，很容易计算出函数 $f(x)$的值，而对大多数 y 属于 Y，要确定满足 $y=f(x)$的 x 在计算上比

较困难(假设至少有这样一个 x 存在)。注意,不能将单向函数的概念与数学意义上的不可逆函数的概念混同,因为单向函数可能是一个数学意义上可逆或者一对一的函数,而一个不可逆函数却不一定是单向函数。

目前,还没有人能够从理论上证明单向函数是存在的。单向函数存在性的证明将意味着计算机科学中一个最具挑战性的猜想 $P=NP$,即 NP 完全问题的解决,而关于 NP 完全性的理论却不足以证明单向函数的存在。现实中却存在几个单向函数的“候选”。说他们是“候选”,是因为他们表现出了单向函数的性质,但还没有办法从理论上证明它们一定是单向函数。

一个最简单的、大家熟知的“候选”单向函数就是整数相乘。众所周知,不管给定两个多大的整数,我们很容易计算出它们的乘积,而对于一个 300 位左右的十进制整数,即使已知它是两个大小差不多(150 位左右的十进制数)的素数之积,用世界上计算能力最强的计算机,也没有办法在一个合理的时间内分解出构成这个整数的两个素数因子来。这里讲的“合理的时间”是指一个可度量的相当长的时间,如人类或者地球的寿命等。

显然,单向函数不能直接用作密码体制,因为如果用单向函数对明文进行加密,即使是合法的接收者也不能还原出明文,因为单向函数的逆运算是困难的。与密码体制关系更为密切的概念是陷门单向函数。一个函数 $f: X\to Y$ 称为是陷门单向的,如果该函数及其逆函数的计算都存在有效的算法,而且可以将计算 f 的方法公开,即使由计算 f 的完整方法也不能推导出其逆运算的有效算法。其中,使得双向都能有效计算的秘密信息叫做陷门(trap door)。

需要提醒的是,不能顾名思义地认为陷门单向函数是单向函数。事实上,陷门单向函数不是单向函数,它只是对于那些不知道陷门的人表现出了单向函数的特性。

提出公钥加密的动机是简化密钥分配和管理,实现签名等功能,是当前密码学领域的最大进步。

2.4.2 RSA 公钥密码算法

RSA 密码体制是目前为止最为成功的非对称密码算法,它的安全性是建立在“大数分解和素性检测”这个数论难题的基础上,即将两个大素数相乘在计算上容易实现,而将该乘积分解为两个大素数因子的计算量相当大。虽然它的安全性还未能得到理论证明,但经过 20 多年的密码分析和攻击,迄今仍然被实践证明是安全的。

RSA 使用两个密钥,一个公共密钥,一个私有密钥。如用其中一个加密,则可用另一个解密,密钥长度在 40~2048 位之间可变,加密时也把明文分成块,块的大小可变,但不能超过密钥的长度,RSA 算法把每一块明文转化为与密钥长度相同的密文块。密钥越长,加密效果越好,但加密解密的开销也大,所以要在安全与性能之间折衷考虑,一般 64 位是较合适的。RSA 的一个比较知名的应用是 SSL,在美国和加拿大 SSL 用 128 位 RSA 算法,由于出口限制,在其他地区(包括中国)通用的则是 40 位版本。

RSA 算法研制的最初理念与目标是努力使互联网安全可靠,旨在解决 DES 算法秘密密钥利用公开信道传输分发的难题。而实际结果不但很好地解决了这个难题;还可利

用 RSA 来完成对电文的数字签名以对抗电文的否认与抵赖；同时还可以利用数字签名较容易地发现攻击者对电文的非法篡改，以保护数据信息的完整性。

RSA 算法描述如下：

(1) 密钥生成。选择两个互异的大素数 p 和 q，n 是二者的乘积，即 $n=pq$，使 $\Phi(n)=(p-1)(q-1)$，$\Phi(n)$为欧拉函数。随机选取正整数 e，使其满足 $\gcd(e,\Phi(n))=1$，即 e 和 $\Phi(n)$互质，则将(n,e)作为公钥。

求出正数 d，使其满足 $e\times d=1 \bmod \Phi(n)$，则将$(n,d)$作为私钥。

(2) 加密算法。对于明文 M，由 $C=M^e \bmod n$，得到密文 C。

(3) 解密算法。对于密文 C，由 $M=C^d \bmod n$，得到明文 M。

如果窃密者获得了 n、e 和密文 C，为了破解密文，他必须计算出私钥 d，为此需要先分解 n 为 p 和 q。为了提高破解难度，达到更高的安全性，一般商业应用要求 n 的长度不小于 1024 位，更重要的场合不小于 2048 位。

RSA 算法提出以后，引起了许多密码分析学家的兴趣，提出了一些针对于 RSA 的攻击方法：如对 RSA 的公共模数攻击、对 RSA 的低加密指数攻击、对 RSA 的低解密指数攻击；对 RSA 的选择密文攻击。根据这些成功的攻击，Jadith Moore 列出了使用 RSA 的一些限制：

(1) 知道了对于一个给定模数的一个加/解密密钥指数对，攻击者就能够分解这个模数。

(2) 知道了对于一个给定模数的一个加/解密密钥指数对，使攻击者无须分解 n 就可以计算出别的加/解密对。

(3) 在通信网络中，利用 RSA 的协议不应该使用公共模数。

(4) 消息中应该使用随机数填充以避免对加密指数的攻击。

(5) 解密指数应该大。

属于基于大整数因式分解困难问题的公钥密码体系的公钥密码还包括 Rabin 算法和 Williams 算法。

例 2-9 选择两个大素数 $p=7$，$q=17$，$p\neq q$。计算 $n=pq=7\times17=119$，$\Phi(n)=(p-1)(q-1)=6\times16=96$。96 的因子有 2、3，因此 e 不能有 2 和 3 的因子；

- 选择整数 $e=5$(公钥，即加密密钥)，使 $\gcd(e,\Phi(n))=1$。
- 选择整数 $d=77$(私钥，即解密密钥)，使 $d\cdot e \bmod \Phi(n)=1$，$(5\times77) \bmod 96=385 \bmod 96=1$。
- 公钥：$KU=\{e,n\}=\{5,119\}$，私钥：$KR=\{d,n\}=\{77,119\}$。

2.4.3 ElGamal 算法

ElGamal 为目前著名的公开密钥密码系统之一，是由 ElGamal 于 1985 年提出的。ElGamal 密码系统可作为加解密、数字签名等之用，其安全性是建立于离散对数(discrete logarithm)问题之上的，即对于 $y=g^x \bmod p$，给定 g、p 与 y，求 x 为计算上不可行。ElGamal 算法包括密钥生成、加密过程、解密过程。

1. 密钥生成

(1) 任选一个大质数 p,使得 $p-1$ 有大质因数。

(2) 任选一个 mod p 之原根 g。

(3) 公布 p 与 g。

使用者任选一私钥 $x \in Z_p$,并计算密钥 $y = g^x \bmod p$。

2. 加密过程

(1) 任选一个数 $r \in Z_p$ 满足 $\gcd(r, p-1) = 1$,并计算

$$c_1 = g^r \bmod p, \quad c_2 = m \times y^r \bmod p$$

(2) 密文为 $\{c_1, c_2\}$。

3. 解密过程

(1) 计算 $w = (c_1^x)^{-1} \bmod p$。

(2) 计算明文 $m = c_2 \times w \bmod p$。

ElGamal 方法具有以下优点:

① 系统不需要保存秘密参数,所有的系统参数均可公开。

② 同一个明文在不同的时间由相同加密者加密会产生不同的密文(机率式密码系统),但 ElGamal 方法的计算复杂度比 RSA 方法要大。

2.4.4 椭圆曲线公钥密码算法

公开密钥密码学的数学理论早在百年前就已经很完备了,只是随着当前计算机技术的进步,将其应用开发出来,RSA、ElGamal 等密码系统都是如此,而椭圆曲线在代数学与几何学上广泛的研究已超出百年之久,已有丰富且深厚的理论,而椭圆曲线系统第一次应用于密码学是 1985 年由 Koblitz 与 Miller 分别提出,随后有两个较著名的椭圆曲线密码系统被提出;一为利用 ElGamal 的加密法,一为 Menezes-Vanstone 的加密法。以下将介绍椭圆曲线的定义、加法运算与反元素运算。

1. 椭圆曲线的定义

令 $p>3$ 为质数,在 GF(p)中的椭圆曲线 $E: y^2 = x^3 + ax + b \bmod p$,其中,$4a^3 + 27b^2 \neq 0 \pmod p$。曲线上另定义一个无穷远点 O,对任一点 $A \in E$,$A+O=O+A=A$。

2. 加法运算

令 $A=(x_1, y_1)$ 与 $B=(x_2, y_2)$为 E 上的点,则若 $x_2 = x_1$ 且 $y_2 = -y_1$,则 $A+B=O$;否则 $A+B=(x_3, y_3)$,其中:

$$\begin{cases} x_3 = \lambda^2 - x_1 - x_2 \\ y_3 = \lambda(x_1 - x_3) - y_1 \end{cases} \tag{2-22}$$

$$\lambda = \begin{cases} \dfrac{y_2 - y_1}{x_2 - x_1}, & A \neq B \\ \dfrac{3x_1^2 + a}{2y_1}, & A = B \end{cases} \tag{2-23}$$

注意：椭圆曲线运算中，大写参数表示点，小写参数表示数值。

椭圆曲线中的乘法运算是透过加法运算达成的。为了加快速度，可以用倍加的运算来达成。例如，$4P$ 计算时，由于 $4P=2P+2P$，再计算 $2P=P+P$ 即可。

3. 反元素运算

点 $A=(x,y)$的反元素为$-A=-(x,y)=(x,-y)$。

$A+(-A)=(-A)+A=O$，此时 O 称为乘法单位元素。

例 2-10 在椭圆曲线 E：$y^2=x^3+x+6 \pmod{11}$上的点有：

(2,4) (2,7) (3,5) (3,6)

(5,2) (5,9) (7,2) (7,9)

(8,3) (8,8) (10,2) (10,9)

再加上 O 共有 13 点。注意在计算点时，要检验 x^3+x+6 的值是否属于 QR_{11}。除了 O 以外，任意点均可以视为 E 的始元素(primitive element)。

令定义于 Z_p 的椭圆曲线 E 的所有点的个数为$\# E$，则满足：

$$p+1-2\sqrt{p} \leqslant \# E \leqslant p+1+2\sqrt{p} \tag{2-24}$$

4. 椭圆曲线密码体制

设 GF(p)是一个有限域，GF(p)上的椭圆曲线是指满足 Weirstrass 方程：

$$y^2 + a_1xy + a_3y = x^3 + a_2x^2 + a_4x + a_5 \quad (\text{无奇点}, a_1, a_2, a_3, a_4, a_5 \in \text{GF}(p)) \tag{2-25}$$

的所有解(x,y)与无穷远点 O 构成的非空集合。

选取一点 $G \in E(\text{GF}(p))$作为公共基点，要求这个公共基点的阶 $l=\text{ord}\ G$ 是一个素数阶，G 为生成元，$\langle G\rangle$是由点 G 生成的 p 阶循环子群。

对于 $Q=dG, d\in Z_q, G、Q\in E(\text{GF}(p))$，已知 G、Q 求 d 称为 ECDLP(椭圆曲线离散对数问题)。基于椭圆曲线的各种密码体制的安全性最终可归结为解 ECDLP 问题，当数据量足够大以致 ECDLP 问题无法解决时，就认为该密码体制是安全的，具有 160 位数据长度的 ECDLP 问题在目前被认为是安全的。

一般的椭圆曲线密码体制都基于以下运算：

① 存在一个容易计算的函数 $f: m \rightarrow P(m)$。

② 选取整数 e，$1<e<N$，选取整数 d，使得 $de=1 \pmod N$，由 $deP(m)=P(m)$，可恢复 $P(m)$。

③ 选取整数 a，$1<a<N$，由 $P(m)=P(m)+aP-aP$，可恢复出 $P(m)$。

SECG 的标准文档 SEC1 中，对有限域 GF(p)上的椭圆曲线域参数 T 定义为如下六

元组：$T=(p,a,b,G,n,h)$。其中，$a,b\in \mathrm{GF}(p)$，满足方程 $y^2\equiv x^3+ax+b$，$G=(x_G,y_G)$为曲线上的基点，基点 G 的阶 n 为一素数，整数 h 为余因子，$h=\#E(\mathrm{GF}(p))/n$，$\#E(\mathrm{GF}(p))$为椭圆曲线的阶。

一个典型的椭圆曲线公钥密码可以描述如下：

设 p 是不等于 3 的素数，椭圆曲线 $E(\mathrm{GF}(p))$包含一个循环子群 A，在 A 中离散对数问题是难处理的。选取 $\alpha\in E$，$0<\alpha<\#A-1$，计算 $\beta=a\alpha$，将 α、β 值公开作为公钥，保密 a 作为私钥。

1）加密过程

设明文 $m=(m_1,m_2)\in Z_p^*\times Z_p^*$，即 m_1,m_2 均属于 Z_p^*。下面对明文进行加密：

① 选取整数 k，$0<k<\#A-1$，k 保密。

② 计算：$y_0=k\alpha$、$(c_1,c_2)=k\beta$、$y_1=c_1m_1 \pmod q$、$y_2=c_2m_2 \pmod q$。

③ 则密文 $c=(y_0,y_1,y_2)$，将其发送给接收方。

2）解密过程

① 接收方接收到密文 c。

② 计算：$(c_1,c_2)=ay_0$。

③ 通过下列运算恢复明文：$m=(y_1c_1^{-1} \pmod q), y_2c_2^{-1} \pmod q))$。

椭圆曲线是一种能够适应未来通信技术和信息安全技术发展的新型密码体制。对于 q 元有限域上的椭圆曲线，q 为 160 位时，RSA 密码体制需要 1024 位的模数才能达到同等的安全强度。也就是说，椭圆曲线密码体制在相同的安全强度下所要求的密钥强度仅是 RSA 的 1/6，因此在运算速度和存储空间方面具有很大的优势，在实际应用中具有很大的使用价值。

2.5 新型密码技术

2.5.1 新型密码技术简介

由于目前常用的一些常规密码体制还存在着一些缺陷，人们仍在不断研究安全性更高的加密手段。与此同时，现代计算技术的发展，也为破译加密系统提供了强有力的工具，很多现有系统已被成功破译，在这种情况下，采用异于常规密码学的加密基础，研究新型密码技术开始引起许多研究人员的兴趣。

这里，将对两种新型的密码技术混沌密码技术和量子密码技术进行介绍。

2.5.2 混沌密码技术

混沌是确定性系统由于内秉随机性而产生的外在复杂表现，是一种貌似随机的非随机运动。被誉为“混沌之父”的美国科学家 Lorenz 曾经也给出过一个通俗的定义：一个真实的物理系统，在排除了所有的随机性影响以后，仍然有貌似随机的表现，那么这个系

统就是混沌的。Lorenz 的这个定义说出了混沌的如下基本特征：

(1) 混沌是系统固有的。系统所表现出来的复杂性是系统自身的、内在的因素所造成的，并不是在外界的干扰下所产生的，是系统内在随机性的表现。

(2) 混沌是具有确定性的。混沌的确定性分为两个方面，首先，混沌系统是确定的系统，是一个真实的物理系统；其次，混沌的表现是貌似随机，而不是经典意义上的随机，系统每一时刻的状态都受到前一时刻状态的影响，而不是像随机系统那样随意出现，这是和随机系统不相同的。

(3) 混沌系统的表现具有复杂性。混沌系统的表现是貌似随机的，它既不是周期运动的，也不是准周期运动的，混沌系统具有良好的自相关性和低频带宽的特点。

混沌理论中的一些基本概念，如混合性、保测变换以及敏感性被认为是应用于密码学中的非常有用的特性，下面将介绍混沌应用于密码学所具备的一些重要特性，主要包括对初值和系统参数的敏感性、类随机性和不可预测性。

(1) 对初值和系统参数的敏感性。Brown R. 和 Chua L. O. 曾指出：图灵机(数字计算机的一种数学模型)是如此的简单，以至于不可能是复杂性的来源，因此(系统)复杂性必定来自初始条件。

系统对初值的敏感性有一个数学定义，据此可作定量计算，但其计算结果与 Lorenz 意义上的对初值敏感性可能是不一致的。在许多文献中，以相轨按指数律发散(或初值的小误差按指数规律放大)作为对初始值敏感的标准。但是，这个提法是不严格的。容易理解，在应用科学和工程上，采用 Lorenz 意义上的对初值敏感的概念及其利用计算机仿真的检验方法(利用短时间内相轨的急剧分离及其直观性)是可行的。

(2) 类随机性。混沌具有类随机性在学术界是一致肯定的。混沌过程可以由算法来定义，而随机过程则不可以，这是其重要差别。另外，混沌过程的随机性实质上属于内秉随机性。由于这些差别，把混沌的随机性称为类随机性是比较合适的。

掷硬币试验是一种随机试验，属于古典概型的典型之一。由于假定硬币是理想匀质的，若其一面出现记为 1，另一面出现记为 0，则此试验是将整数映射到集合{0,1}的一个映射。这个映射可以用来定义一个序列，因而这种试验的每一次试验可得到一个二进制数。这种试验的随机性可以表述如下：无论投掷试验进行多少次，都不可能写出一个可由以前试验所得的值计算出下一次所得值的公式。混沌的类随机性的含义与这一表述是一致的。也就是说，混沌的类随机性意味着混沌的不可预测性。

(3) 不可预测性。混沌具有不可预测性在许多文献中都有明确的叙述。混沌吸引子起着局部噪声放大器的作用，一个小的起伏会导致相轨迅速产生很大的偏离；过去和将来(系统状态)没有必然的联系。根据上面的随机性的表述，可以给出定义：对于动态系统的一个变量和任意给定的时间 $t_0>0$，如果总可以找到不大的时间间隔 $\Delta t_0>0$，而不可能找到这样的一个通用公式或算法：它可以用来进行由 t_0 时的变量值确定 $t_0+\Delta t_0$ 时的值的计算，则该系统是不可预测的。由此定义可知，利用数字计算机对混沌系统进行仿真时由一个初始所得的相轨不是该系统的解。换句话说，此时计算机及其算法所构成的系统不是原混沌系统的准确模型，问题在于丧失了初始值的随机性。然而，利用数字计

算机对混沌系统进行统计分析所得结果能够近似反映混沌系统的统计特性。

L. Kocarev 指出了混沌系统和常规密码学的关系，如图 2-9 所示。混沌密码学和常规密码学具有一些相同点：对系统参数和初始值敏感、类随机特性、长期的不稳定非周期轨道；密码算法的加密轮数导致扩散和混淆效果的产生，混沌密码中混沌系统的迭代次数将系统从初始区域遍布至整个相空间；混沌系统的系统参数可以作为密码系统的密钥。混沌密码学与常规密码学不同之处在于常规密码学是定义在有限整数集上，而混沌密码学中混沌系统只在实数集上有意义。与此同时，在混沌密码领域，还没有建立和常规密码学相类似的性能和安全性分析理论。



图 2-9 混沌系统和常规密码学的比较

按照常规密码学的分类方法，类似地，可以将混沌密码分为混沌对称密钥密码和混沌公钥密码，其中混沌对称密钥密码又可以分为混沌流密码和混沌分组密码。

1. 混沌对称密钥密码

一般来说，混沌对称密钥密码有两种通用的设计方法：第一种是使用一个或多个混沌系统生成伪随机密钥流，然后使用该密钥流对明文进行掩盖加密；第二种是使用明文/密钥作为混沌系统的初始值/控制参数，通过多次迭代/反向迭代运算得到明文。第一种设计方法对应常规密码学中的流密码。在此称为混沌流密码；第二种设计方法对应常规密码学中的分组密码。类似地，我们称之为混沌分组密码。

1）典型的混沌流密码

典型的混沌流密码主要包括基于混沌伪随机数发生器的流密码以及基于混沌逆系统方法设计的流密码。

由于混沌系统轨道的不可预测性，很多研究集中在使用混沌系统构造伪随机数发生器（Peusodo-Random Number Generator，PRNG）上。基于混沌伪随机数发生器的流密码的核心部分是混沌伪随机数发生器，以混沌伪随机数发生器的输出作为密钥流对明文进行掩盖加密。从笔者目前所掌握的文献来看，主要有以下两种生成混沌伪随机数的方法（见图 2-10）：

① 抽取混沌轨道的部分或全部的二进制比特。

② 将混沌系统的定义区间分割成为 n 个互不相交的子区间，每个区域用唯一的 0～$n-1$ 的数值标记。通过迭代混沌系统，看迭代变量值落入哪个子区间，获得相应子区间的标记，从而得到相应的伪随机数。

这两种混沌伪随机数的生成方法之间存在一定的关系：即从数字的角度来看，方法一中从混沌轨道的部分或全部的二进制比特所组成的数值也是在某一个范围的，故可看做是方法二的一个特例；同时，当方法一抽取混沌轨道全部的二进制比特时，方法二则可以看做是方法一的一个特例。

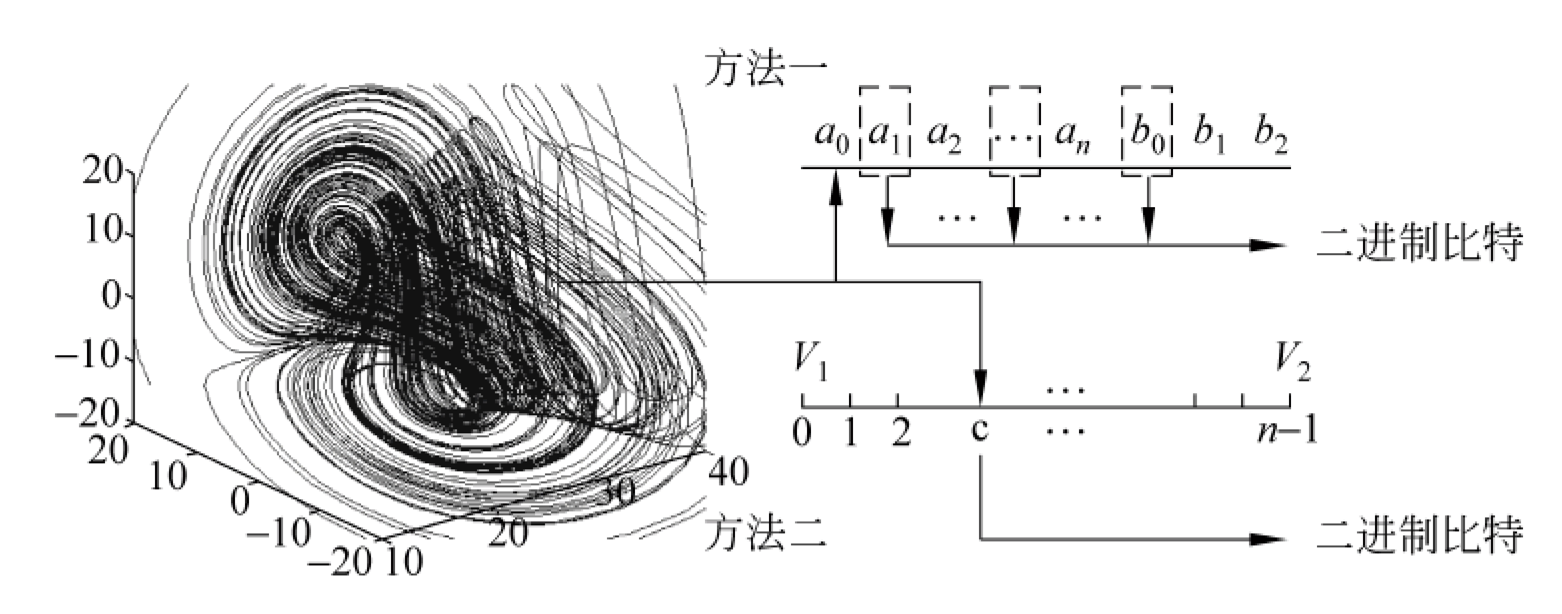


图 2-10 混沌伪随机数生成方法

常见的几种基于混沌逆系统的混沌密码的基本结构都可以表示如下：$y(t)=u(t)+f_e(y(t-1),\cdots,y(t-k)) \bmod 1$，其中 $u(t)$、$y(t)$ 分别表示明文和密文；$f_e(\cdot)$ 是一个从反馈密文生成掩盖明文的伪随机密钥流的 k 元函数。在不同的方法中，$f_e(t)$ 的定义不同。如 $f_e(t)=ay(t-1)+by(t-2)$，$f_e(t)=F^m(y(t-1),p)$ 等。其中，$F(x,p)$ 是一个在 $L-\text{bit}(L<m)$ 有限精度下实现的逐段线性混沌映射：

$$f(x,p)=\begin{cases}x/p, & x\in[0,p)\\(x-p)/(0.5-p), & x\in[p,0.5]\\F(1-x,p), & x\in[0.5,1)\end{cases} \tag{2-26}$$

2）典型的混沌分组密码

典型的混沌分组密码包括：基于逆向迭代混沌系统的分组密码，基于正向迭代混沌系统的分组密码和基于混沌 S 盒(Substitution-Box)的分组密码。

T. Habutsu 等使用逆向迭代混沌系统构造密码系统。其基本方法是：给定一个秘密密钥 p 和如下的 tent 映射 $F_p(x)$ 以及它的逆向映射 $F_p^{-1}(x)$：

$$F_p(x)=\begin{cases}x/p, & x\in[0,p]\\(1-x)/(1-p), & x\in(p,1]\end{cases} \tag{2-27}$$

$$F_p^{-1}(x)=\begin{cases}px, & b=0\\1-(1-p)x, & b=1\end{cases} \tag{2-28}$$

这里 b 是一个在集合{0,1}中均匀分布的随机变量。该密码系统按照如下方式加密每个明文分组 $P\in(0,1)$，得到密文 C：$C=F_p^{-n}(P)$，这里需要生成 n 个随机比特 $b_1\sim b_n$ 来确定每次逆向迭代的输出。在解密过程中，明文 P 则通过正向迭代进行恢复：

$$P=F_p^n(C)=F_p^n(F_p^{-n}(P))$$

L. Kocarev 等提出了一种类似 Feistel 网络结构的混沌分组密码方案，其结构如图 2-11 所示。设 B_0 为 64 位的明文分组，$x_{i,0},x_{i,1},\cdots,x_{i,7}$ 表示分组 B_i 的 8 位，也就是说 $B_i=x_{i,0},x_{i,1},\cdots,x_{i,7}$。加密过程包含 r 轮对明文分组相同的变换处理，每一轮加密过程表示如下：

$$x_{i,k+1}=x_{i-1,k}\oplus f_{k-1}[x_{i-1,1},\cdots,x_{i-1,k-1},z_{i-1,k-1}] \tag{2-29}$$

其中，$i=1,2,\cdots,r$，$k=1,2,\cdots,8$，$f_0=z_{i,0}$，$x_8\equiv x_0$，$x_9\equiv x_1$，$z_{i,0},\cdots,z_{i,7}$ 为控制第 i 轮运算的子密钥 z_i 的 8 个比特。函数 $f_1,f_2,\cdots,f_7$ 为：

$$f_j=f(x_1,x_2,\cdots,x_j,z_j) \tag{2-30}$$

其中，$j=1,2,\cdots,7$，$f:M\to M$，$M=\{0,2,\cdots,255\}$为一个离散化的混沌映射。输出分组 $B_i=x_{i,0},x_{i,1},\cdots,x_{i,7}$ 作为下一轮的输入分组，$B_r=x_{r,0},x_{r,1},\cdots,x_{r,7}$ 为密文分组。其解密过程为对密文分组 B_r 进行 r 轮解密操作，得到明文 B_0。每一轮解密过程为：

$$x_{i-1,k}=x_{i,k+1}\oplus f_{k-1}[x_{i-1,1},\cdots,x_{i-1,k-1},z_{i-1,k-1}] \tag{2-31}$$

其中，$k=1,2,\cdots,8$，$f_0=z_0$，$x_8\equiv x_0$，$x_9\equiv x_1$。通过对以指数函数和 Logistic 混沌映射为例对密码方案的性能进行了评估，表明具有可接受的线性逼近概率和差分逼近概率，具有较好的抵抗线性攻击和差分攻击的能力。

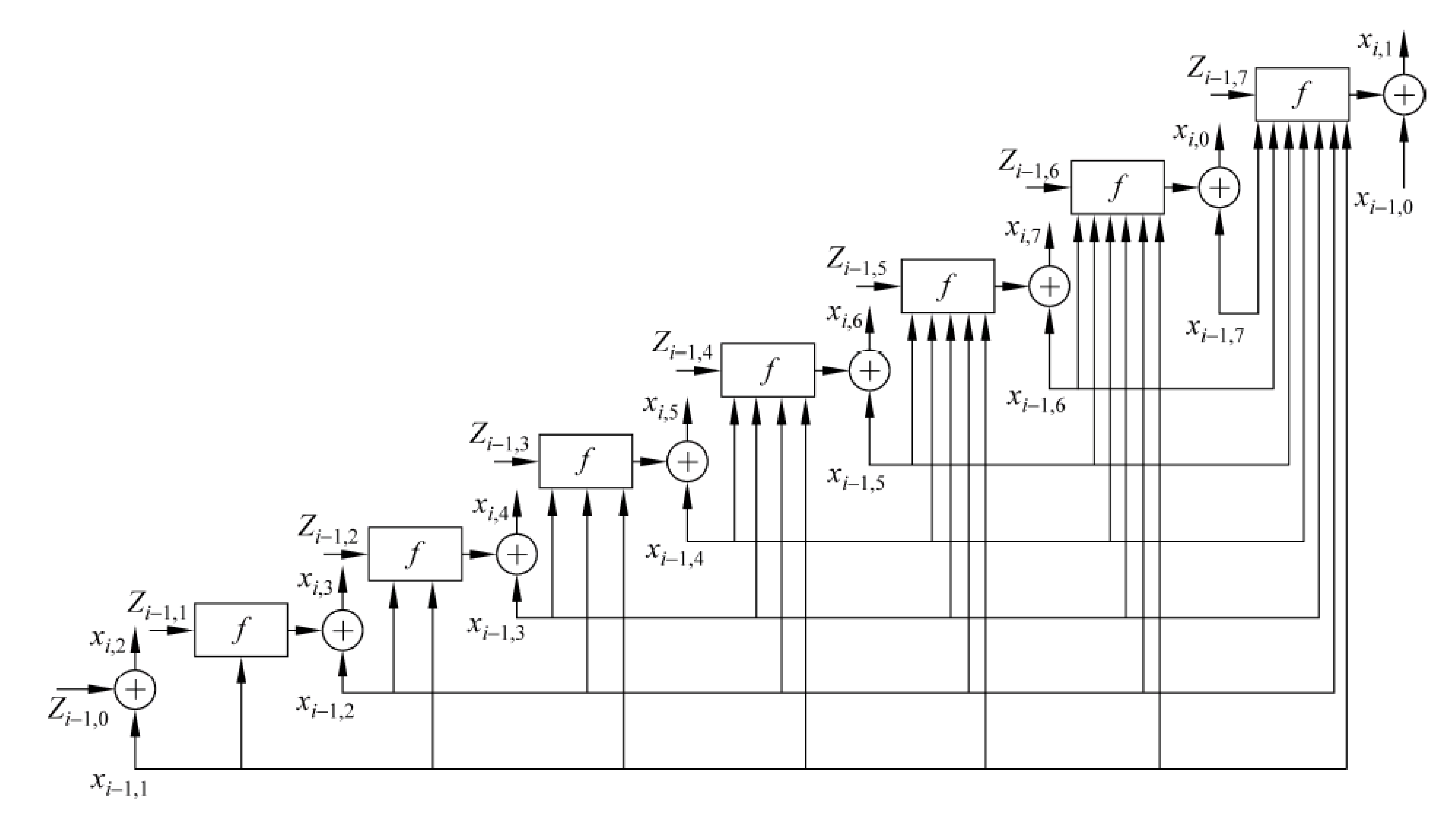


图 2-11　一种类似 Feistel 网络的混沌分组密码结构图

2. 混沌公钥密码

根据目前所收集到的资料，一共只有 6 种混沌公钥密码系统被公开报道过。其中两种混沌公钥密码是基于胞元自动机的，虽然目前还未有相关的密码分析报道，但这两种公钥密码的安全性还有待考察。这里将简单介绍其他 4 种混沌公钥密码。

Fengi Hwu 提出了一类混沌公钥密码。实际上，该密码是 ElGamal 公钥方案的一个变形，其基本工作原理如下：每个用户选择并公开一组参数(a_0,a_n,α)作为公开密钥同时选择一个整数 n 作为私钥，这里 α 在$\{1,2,\cdots,p-1\}$上均匀分布，a_n 迭代下述混沌映射 n 次(以 a_0 为初始值)得到：$F(x)=\alpha x \bmod p$ 或者 $F(x)=x^2 \bmod p$(或者其他更加复杂的数字化混沌映射)。整数 p 是一个大素数(200 比特左右)并且使得 $p-1$ 有一个大的素因子，α 是 p 的一个生成元。然后以类似 ElGamal 公钥方案的方式进行加密和解密。

Tenny Roy 等提出了一种基于 DDE 的混沌公钥密码系统。该类公钥密码系统的工作原理为：将一个 D_T+D_R 维的动力学系统分割为一个包含 D_T 系统变量的发送方子系统(公开)和包含 D_R 维系统变量的接收方(秘密)子系统。发送方传送一个嵌入了明文信号 $m(n)$的标量信号 $s_t(n)$给接收方，接收方则反送另外一个标量信号 $s_r(n)$给发送方。给定整个动力学系统的两个不同的吸引子，每个明文比特 $m(n)$通过判断系统在 L 次混沌

迭代后收敛到哪个混沌吸引子来得到。该 DDE 系统的性能通过一个耦合映射网格得到了演示。在该混沌公钥密码系统中,对每个比特而言,上述的混沌吸引子需要经常改变以抵抗明文攻击,但这种吸引子的改变可能会加大接收方的运算开销。算法的提出者分析了当信道中存在噪声的情况下如何抵抗一类基于隐 Markov 模式(Hidden Markov Model,HMM)的攻击,并指出在噪声较大的情况下可能会导致安全性降低。

L. Kocarev 等人提出了一种基于 p 次 Chebyshev 多项式映射混沌在 $T_0=0$ 和 $T_1=x$ 时的公钥密码方案,该方案受 RSA 和 Rabin 算法思想 $X_{n+1}=(X_n)^p(\bmod N)$的启发,利用 Chebyshev 映射的半群特性,即

$$T_r(T_s(x)) = T_s(T_r(x)) = T_{r\cdot s}(x) \tag{2-32}$$

构造了一个几乎与 RSA 算法几乎一模一样的公钥算法。虽然该方案在实数域具有运算速度快的优点,但同时指出了该公钥密码在已知密文攻击下是不安全的。

最新出现的混沌公钥密码方案是 2005 年由国立新加坡大学的学者 Wang Xingang 和 Gong Xiaofeng 在 *CHAOS* 发表的基于耦合映射格子通用同步(GSCML)的公钥密码方案,它以 Merkle 难题(Merkle's puzzles)和预测通用同步函数的困难性为密码系统的安全基础。该方案由一个新设计的单向耦合映射格子系统(一种混沌动力系统)和 Merkle 难题相结合构成。分析和讨论表明该公钥密码算法将混沌动力系统与 Merkle 难题相结合具有如下新的特点:它能够满足不同安全要求变化的需要,密钥的推导过程是可管理的,且能够推广到实际应用。

2.5.3 量子密码技术

量子密码学是当代密码理论研究的一个新领域,近年来在密码理论研究中逐渐热门起来。量子密码学的思想最早是由 20 世纪 60 年代末美国人 Stepphen Wiesener 在一份手稿中首先提出的,后来美国 IBM 公司 Thomas. Waston 研究中心的 Charles H. Bennett 与加拿大蒙特利尔大学的 GillesBrassard 受其思想影响在 1982 年美洲密码学会上发表了第一篇论文,1984 年提出了量子密码协议,现在被通称为 BB84 协议,并于 1989 年制作了一台原型样机。后来,英国防卫研究署、瑞士日内瓦大学、英国电信实验室和美国国家实验室分别进行了类似的研究,用相位编码的方式实现了 bb84. QKD 方案,光线传输长度达到了 10 千米。到 1995 年,在光纤中的传输距离达到了 30 千米。2000 年,美国洛斯阿拉莫(LosAlmos)国家实验室在自由空间里进行的量子密钥分配的传输距离达到了 1.6 千米。2003 年,欧洲小组在自由空间中的距离达到了 23 千米。目前他们正在为地面与低轨道卫星之间的量子密码通信试验做准备。2006 年,中国科学技术大学潘建伟教授领导的研究小组,在国际上首次成功地实现了两粒子复合系统量子态的隐形传输,并且第一次成功地实现了对六光子纠缠态的操纵。

目前,量子秘密共享、量子签名和量子认证都是最近发展起来的量子密码技术研究方向。量子密码是以 Heisenberg 测不准原理(光子的偏振现象)和 EPR 效应为物理基础,利用光纤异地产生物理噪声。它可以真正地实现一次一密密码,构成理论上不可破译的密码体制。光子不能被克隆的性质使量子密码编码操作过程不能被完全窃听,一旦

存在窃听也可以察觉，并可以设法消除。

1. 量子密钥产生与分发的物理基础

(1) 光子的偏振现象：每个光子都有一个偏振方向即电场的振荡方向。在量子密码学中用到两种光子偏振，即线偏振和圆偏振，其中，线偏振可取两个方向：水平方向和垂直方向；圆偏振包括左旋和右旋两种情况。在量子力学中，光子的线偏振状态和圆偏振状态是一对共轭可观测量，也就是说，光子的线偏振态状态与圆偏振态状态是不可同时测量的。值得说明的是，在同一种偏振态下的两个不同的方向是可完全区分的，例如在线偏振态中的水平方向和垂直方向是可完全区分的，因而可同时准确测量。

(2) Heisenberg 测不准原理：光子的一对共轭偏振态是互补的，正是这一本质特征为 BB84 协议提供了实现的基础。实际上，在量子力学中任何两组不可同时测量的物理量都是共轭的，都满足互补性，在进行测量时，对其中一组量的精确测量必然导致另一组量的完全不确定，即遵循量子力学的基本原理——Heisenberg 测不准原理。

(3) EPR(Einstein Podolsky Rosen)纠缠效应：一个球对称原子系统中，同时向两个相反的方向发射两个相干光子，初始时这两个光子都是未被极化的，测量其极化态(偏振态)时，对两个光子中的任一个进行测量可得到测量光子的极化态，同时另一个光子的极化态也被同时确定，但两个光子的极化态的方向相反。

(4) 单量子不可克隆定理：所谓"克隆"是指原来的量子态在不被改变的情况下，在另一个系统中产生一个完全相同的量子态。对于一个未知的单量子态不能被完全拷贝。对两个非正交的量子态不能被完全拷贝。要从编码在非正交量子态中获得信息，不扰动这些态是不可能的。

2. 量子密钥产生与分发的实现过程

量子密码学还不能像对称、公钥加密体制那样能对数据直接进行加密处理，目前只能进行安全密钥分发。量子密钥产生与分发的实现过程大致可分为 5 个过程。

(1) 量子传输：不同的协议有不同的量子传输方式，但有一个共同点：它们都利用量子力学原理或量子现象来实现。在实际的通信中，光子态序列中光子的极化态将受到噪声和 Eve(窃听者)的影响，但按照 Heisenberg 测不准原理，Eve 的干扰必将导致光子极化态的改变，这必然会影响 Bob 的测量结果。由此可对 Eve 的行为进行判定和检测。

(2) 数据筛选：在量子传输中由于噪声和 Eve 的作用，光子态序列中光子的极化态会发生改变。另外，实际系统中，Bob(信息接收者)的接收仪器不可能有百分之百的正确的测量结果，所有那些在传送过程中没有收到或测量失误，或由于各种因素的影响而不合要求的测量结果，由 Alice(信息发送者)和 Bob 经过比较测量基矢后全部放弃，并计算错误率。若错误率超过一定的阈值，Alice 和 Bob 放弃所有的数据并重新开始，如果是一个可以接收的结果，则 Alice 和 Bob 将筛选后的数据保存下来，所获得数据称为筛选数据(sifted data)。

(3) 数据纠错：所得到的 n 比特筛选数据并不能保证 Alice 和 Bob 各自保存的安全一致性，这可以由各种因素造成，解决这一问题的办法是对原数据进行纠错，如采用奇偶

校验等。这样做的目的是为了减少 Eve 所获得的密钥信息。

(4) 保密加强：保密加强是为了进一步提高所得密钥的安全性而采取的措施(非量子的方法)，其具体实现为：假设 Alice 发给 Bob 一个 n 比特串，Eve 获得的比特为 $t<n$。为了使 Eve 所获得的信息无用，Alice 和 Bob 采用秘密加强技术：公开选取一个压缩函数 $G:\{0,1\}^n \rightarrow \{0,1\}^r$，其中 r 是被压缩后密钥的长度，这样使得 Eve 从 W 中获取的信息和她的关于函数 G 的信息给出她对新密钥 $K=G(W)$ 尽可能少的信息。对任意的 $s<n-t$，Alice 和 Bob 可得到长度为 $r=n-t-s$ 比特的密钥 $K=G(W)$，而 Eve 所获得的信息随 s 按指数减少 $V=f(e-as)$。

(5) 身份确认：以上是假定 Alice 和 Bob 都是合法的，然而在实际通信中，存在 Alice 和 Bob 假冒的情况，为此应在量子密钥的获取过程中加上身份确认这一非量子过程，可采用以往的身份认证方案，亦可从获得的量子密钥中获取认证密钥而实现。后一种方案是从所获得的量子密钥(称为原密钥)中截取一部分作为认证密钥，然后 Alice 和 Bob 用认证密钥进行身份认证。

由于采用的 4 个偏振态光子中线偏振和圆偏振是不对称的，因此它们不可以同时准确测量。由于 Eve 事先不知道这些光子态，Eve 不可能正确地选取每一个光子态的测量基，因此 Eve 测量时，由 Heisenberg 测不准原理可知，会对 Alice 发送的光子态有扰动，这给 Alice 和 Bob 的测量结果中留下痕迹，这样使得 Eve 的目的不可能实现。

3. 量子密码基本协议

在量子密码学中，通信双方的秘密通信是通过量子密钥分配协议的支撑来实现的。在某一加密系统中，依据协议，通信双方能在一个即将作为密钥的秘密比特串问题上达成一致意见。目前，量子密码的协议主要有三种。

1) BB84 协议

BB84 协议是基于两种共轭基的四态方案，其原理是利用单光子量子信道中的测不准原理。Alice 每隔一定时间随机地从 4 个光子极化态 0、$\pi/4$、$\pi/2$、$3\pi/4$ 中任意选取一个发送给 Bob，形成具有一定极化态的光子态序列，并记录每一个光子态对应的基矢类型。Bob 接到 Alice 发送的信号后，开始接收 Alice 发送的光子态序列，Bob 为每一个光子从两种测量基矢中随机地选取一种进行测量，然后记录测量的结果并秘密保存。Bob 接收并测量完 Alice 发送来的极化态光子序列后，向 Alice 公开其测量过程中所用的基矢或测量类型。Alice 进行比较并告诉 Bob 其比较的结果：告诉 Bob 哪些是正确的，哪些是错误的。根据比较结果，Alice 与 Bob 按照事先的约定将经过比较后的所有正确的光子极化态翻译成二进制比特串，从而获得所需的密钥。

2) B92 协议

B92 协议是基于两个非正交态的两态方案，其原理是利用非正交量子态不可区分原理，这是由测不准原理决定的。首先，选择光子的任何两套共轭的测量基，取偏振方向为 0 和 $\pi/2$，$\pi/4$ 和 $3\pi/4$ 的两套线偏振态，并定义 0 和 $3\pi/4$ 代表量子比特 0、$\pi/4$ 和 $\pi/2$ 代表量子比特 1。合法用户 Alice 随机发射偏振态(这里取 0 和 $\pi/4$)，Bob 随机使用偏振态(这里取 $\pi/2$ 和 $3\pi/4$)进行同步测量。下面给出建立密码本的具体步骤：

① Alice 以 0 或 π/4 光子线偏振态随机向 Bob 发射选定的光子脉冲。

② Bob 随机选取 π/2 或 3π/4 方向的检偏基检测，当 Bob 的检测方向与 Alice 所选方向垂直，探测器完全接收不到光子；当成 π/4 时，则有 50％的概率接受到光子。一旦 Bob 测到光子，Bob 就可推测出 Alice 发出的光子的偏振态。

③ Bob 通过公共信道告诉 Alice 所接收到光子的情况，但不公布测量基，并且双方放弃没有测量到的数据（空格表示未接收到光子）；此时如无窃听或干扰，Alice 和 Bob 双方则共同拥有一套相同的随机数序列。

④ Bob 再把接收到的光子转化为量子比特串。

⑤ Bob 随便公布某些比特，供 Alice 确定有无错误。

⑥ 经 Alice 确认无误断定无人窃听后，剩下的比特串就可留下建立为密码本。

这种方法比 BB84 协议简单，但代价是传输速率减少一半，因为只有 25％的光子被接收到。

3）E91 协议

E91 协议是基于 EPR 纠缠对编码实现的，由 Ekert 于 1991 年提出，原理是利用 EPR 效应。其通信过程是：

① 由 EPR 源产生的光子对分别朝 ±Z 方向发送到合法用户 Alice 和 Bob，Alice 任意选择检偏基（线偏振基或圆偏振基）测量接收到的其中一个光子 1，测量的结果由 EPR 关联决定。

② 同时 Bob 也随机用检偏基测量接收到的 EPR 关联对的另一个光子 2，并记录测量结果。

③ Bob 通过公共信道公开其使用的测量基（但不公布测量结果），Alice 告诉 Bob 哪些检偏基选对了，然后双方保留正确的结果并将它转化为量子比特串，再通过商定建立为密码本。它与 BB84 不同的是检验双方保留的数据是用 Bell 不等式检验，如果违反不等式，表明量子信道是安全的，没有被窃听；如果满足不等式时，表明信道有问题即存在窃听者。其安全性源于 Bell 原理，根据量子力学原理该协议是安全的。

思　考　题

2.1　假设使用的密码是移动了 n 位的简单代替密码，试从下面的明文中找出明文和密钥：

CSYEVIXIVQMREXIH

2.2　Playfair 密码可用的密钥有多少个？要求写成接近 2 的乘方的形式。

2.3　用 Hill 加密消息 meet me at the usual place at then rather eight oclock，密钥为 $\begin{bmatrix} 9 & 4 \\ 5 & 7 \end{bmatrix}$。要求写出计算过程和结果，并写出从密文恢复为明文所做的解密算法。

2.4　当海军上尉 John F. Kennedy 下令击沉美国巡逻号 PT-109 时，在澳大利亚的无线站截获了一条用 Playfair 密码加密的消息：

KXJEY　UREBE　ZWEHE　WRYTU　HEYFS
KREHE　GOYFI　WTTTU　OLKSY　CAJPO

BOTEI	ZONTX	BYBWT	GONEY	CUZWR
GDSON	SXBOU	YWRHE	BAAHY	USEDQ

密钥为 royal new Zealand navy。请解密这条消息。

2.5 本题探讨 Vigenère 密码的一次一密版本的用途。在这种方案中，密钥是 0～26 之间的随机数流。例如，如果密钥是 3 19 5…，则密文的首个字母使用 3 个字母的移位加密，第二个字母使用 19 个字母的移位加密，第三个字母使用 5 个字母的移位加密，依此类推。

(1) 使用密钥流 9 0 1 7 23 15 21 14 11 11 2 8 9 加密明文 sendmoremoney。

(2) 使用(1)中产生的密文找到一个密钥，以便该密文解密为 cashnotneeded。

2.6 实现 RC4 算法。假设密钥由下列七个字节构成：

key=(0x1A,0x2B,0x3C,0x4D,0x5E,0x6F,0x77)

(1) 列出初始化阶段之后的 S。

(2) 列出生成 100 字节的密钥流之后的置换 S。

(3) 列出生成 1000 字节的密钥流之后的置换 S。

2.7 解决密钥分配问题的一个办法是使用收发双方都有的一本书的某行字。至少在某些侦探小说中经常把一本书的第一句话作为密钥，这里就从一本富于悬念的侦探小说——Ruth Rendell 的《与陌生人的谈话》中找到一个例子。请不要找到这本书之后再来做这道题。

给定下列消息：

SIDKHKDM AF HCRKIABIE SHIMC LFEAILA

这段密钥是用《沉默的背后》艺术的第一句话和单表代换方法产生的，这句话是：

This is lay thick on steps and the snowflaks driven by the wind looked in the headlights of the cars.

请回答：

(1) 加密算法是什么样的？

(2) 它的安全性怎么样？

(3) 为了使密钥问题简单化，通信双方都同意使用一本书的第一句话或者最后一句话作为密钥，要想改变密钥，他们只需更换一本书就行了。使用第一句话比使用最后一句话要好，这是为什么？

2.8 令 $K=(k_0k_1k_2\cdots k_{55})$ 为 56 位的 DES 密钥。列出 DES 每一轮的 48 位子密钥 k_1，k_2，…，k_{16}。对于每位 k_i 在密钥中使用的次数做成表。能够设计出一种 DES 子密钥扩展算法是的每个密钥位的使用此时都相同吗？

2.9 Alice 的 RSA 公钥是(N,e)=(33,3)，私钥是 d=7。

(1) 如果 Bob 加密消息 M=19 给 Alice，密文 C 是什么？展示 Alice 将 C 解密到 M 的过程。

(2) 令 S 是 Alice 对于消息 M=25 的数字签名结果。S 是多少？如果 Bob 收到 M 和 S，解释 Bob 验证签名的过程。假设这一签名过程验证成功。

2.10 对于椭圆曲线：

$$E: y^2 = x^3 + 11x + 19 \pmod{167}$$

和曲线 E 上的点 $p=(2,7)$，假设使用 E 和 P 进行 ECC 的 Diffie-Hellman 密钥交换，这里 Alice 选择秘密值 $A=12$，Bob 选择秘密值 $B=31$，Alice 发送给 Bob 的值是什么？Bob 发送给 Alice 的值是什么？共享秘密是什么？

2.11 在 ElGamal 数字签名方案中，公钥是三元组(y,p,g)，私钥是 x，这里满足

$$y = g^x \bmod p \qquad (\text{Ⅰ})$$

要对消息 M 签名，选择一个随机数 k，并且 k 与 $p-1$ 互素，计算：

$$a = g^x \bmod p$$

找到值 S 满足

$$M = xa + ks \bmod (p-1)$$

这使用欧几里得算法很容易计算，签名的验证是通过判断

$$y^a a^s = g^M \bmod p \qquad (\text{Ⅱ})$$

(1) 选择(y,p,g)和 x 的值，使得满足式(Ⅰ)，选择明文 M，计算签名，并验证式(Ⅱ)。

(2) 证明该 ElGamal 签名体制的正确性，即证明式Ⅱ总是成立。

提示：使用费马定理：如果 p 是素数，且 p 不能整除 z，那么 $z^{p-1}=1 \bmod p$。

参 考 文 献

[1] 陈鲁生，沈世镒. 现代密码学. 北京：科学出版社，2002.

[2] Stinson D R. Cryptography (Theory and Practice). CRC Press, Boca Raton, 1995.

[3] Denning D E R. Cryptography and Data Security, London: Addison-Wesley Puplishing Company, 1982.

[4] Saloma A. Public-key Cryptography, Springer, 1990.

[5] Douglas R Slinson. 密码学原理与实践，北京：电子工业出版社，2003.

[6] 冯登国，裴定一. 密码学导引. 北京：科学出版社，1999.

[7] Gardner P. Marking. Breaking Codes: An Introduction to Cryptology. Prentice Hall, 2001.

[8] Katzenbeisser S. Information Hiding Techniques for Steganography and Digital Watermarking. Boston: Artech House, 2000.

[9] Singh S. The Code Book: The Science of Secrecy from Ancient Egypt to Quantum Cryptography. New York: Anchor Books, 1999.

[10] Nichols R K. ICSA Guide to Cryptography. New York: McGraw-Hill, 1999.

[11] Fernandes A. Elliptic Curev Cryptography. Dr. Dobb's Journal, 2001: 56-63.

[12] 彭飞. 混沌密码算法及其在安全电子邮件中的应用. 华南理工大学博士学位论文，2006.

[13] 曾贵华. 量子密码学. 北京：科学出版社，2006.

第 3 章 chapter 3

消息认证与数字签名

本章学习目标

消息认证与数字签名是判断数字内容完整性的重要技术。本章将介绍消息认证与数字签名的基本概念、消息认证的模式与认证方式、单向 Hash 函数与消息认证码的基本原理、常用的数字签名及一些认证的方法和技术。

通过本章的学习,应掌握以下内容:

(1) 消息认证的概念、作用及其基本原理。

(2) 单向 Hash 函数与消息认证码的基本概念和数字签名的基本原理。

(3) 认证模式与认证方式,常用的数字签名体制。

3.1 消息认证与数字签名概述

信息安全通常包括两个方面的内容:一方面是保证其保密性,防止通信中的机密信息被窃取或破译,防止发生针对系统的被动攻击;另一方面是保证信息的完整性、有效性,即要确认信息在传输过程中是否被篡改、伪装和抵赖,以及与之通信的对方身份的真实性,防止发生针对系统的主动攻击。

认证是防止主动攻击(如篡改、伪造信息等)的一项重要技术,可用于开放环境中各种信息系统安全性的保护。认证的目的包括以下两个方面:一是验证信息的完整性以及数据在传输或存储过程中是否被篡改、重放或延迟等;一是验证信息发送者的身份是合法的,不是冒充的。

通常,认证和保密的关系是相对独立的,即一个认证系统它不能自动地提供保密的功能,而一个保密系统也不会自然地提供认证的功能。图 3-1 给出的是一个纯认证系统的模型。

在这个系统中发送者通过一个公开信道将信息传送给接收者,接收者除收到消息本身以外,还要通过认证编码器和认证译码器验证消息是否被篡改以及消息是否来自合法的发送者。系统的串扰者是指可截获和分析信道中传送的密文,而且可伪造密文送给接收者进行欺诈的主动攻击者。

消息认证是指通过对消息或者与消息有关的信息进行加密或签名变换进行的认证,

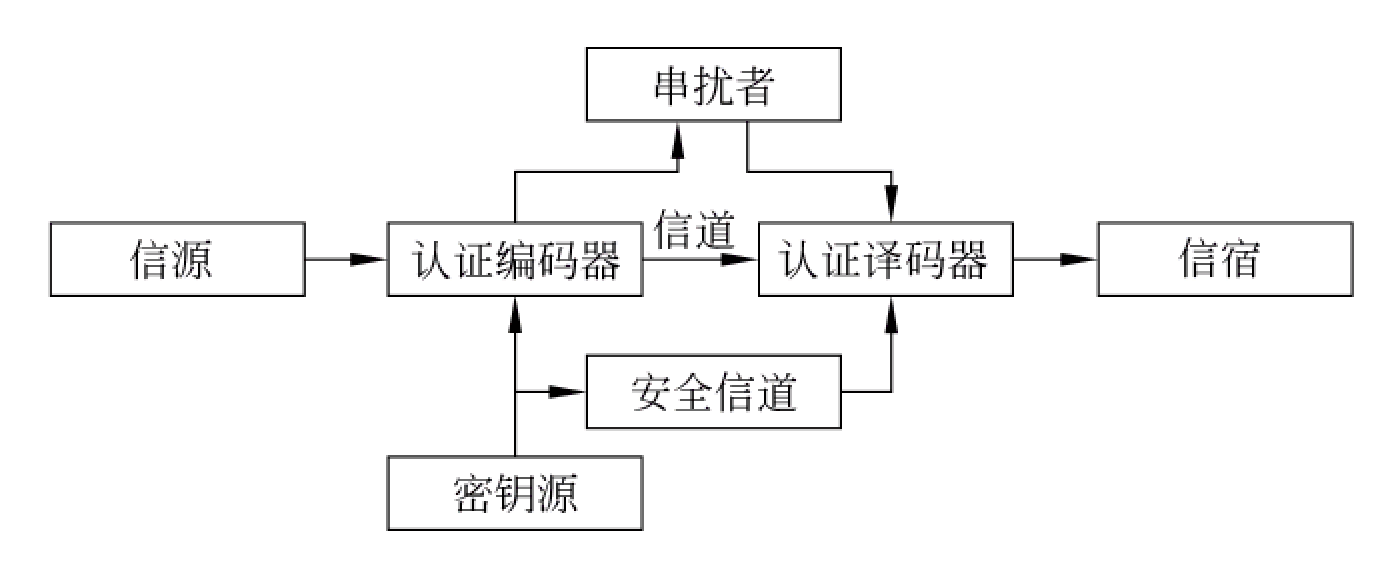


图 3-1　一个纯认证系统的模型

目的是为了防止传输和存储的消息被有意无意地篡改，包括消息内容认证（即消息完整性认证）、消息的源和宿认证（即身份认证）以及消息的序号和操作时间认证等。它在票据防伪中具有重要应用（如税务的金税系统和银行的支付密码器）。

数字签名（又称公钥数字签名、电子签章）是一种类似写在纸上的普通的物理签名，但是它使用了公钥加密技术实现。一套数字签名通常定义两种互补的运算：一种用于签名，另一种用于验证。

3.2　单向 Hash 函数

3.1.1　基本概念

Hash 函数长期以来一直在计算机科学中使用，无论从数学上或别的角度看，Hash 函数就是把可变长度的输入串转换成固定长度的输出串（叫做 Hash 值）的一种函数。

Hash 函数具备以下性质：

(1) H 可适用于任意长度的输入数据块，产生固定长度的 Hash 值。

(2) 对于每一个给定输入数据 M，都能很容易计算出它的 Hash 值 $H(M)$。

(3) 如果给定 Hash 值 h，要逆推出输入数据 M 在计算上不可行，即 Hash 函数具备单向性。

(4) 对于给定的消息 M_1 和其 Hash 值 $H(M_1)$，找到满足 $M_2 \neq M_1$，且 $H(M_2)=H(M_1)$ 的 M_2 在计算上是不可行的，即抗弱碰撞（Collision）性。

(5) 要找到任何满足 $H(M_1)=H(M_2)$ 且 $M_1 \neq M_2$ 的消息对 (M_1, M_2) 在计算上是不可行的，即抗强碰撞性。

这里所说的碰撞，是指如果有两个不同的消息，它们生成的 Hash 值相同，则称发生了一次碰撞。特别需要注意的是，Hash 函数并不提供机密性，且无需使用密钥就可以生成 Hash 值，它非常适合于消息认证。

安全单向 Hash 函数的一般结构如图 3-2 所示。

由图 3-2 可知，单向 Hash 函数重复使用一个压缩函数 f 来实现 Hash 值的生成。压缩函数通常有两个输入，一个是前一阶段的输出 H_{i-1}，另一个来源于消息分组 M_i，最后产生一个输出 H_i，可表达为：

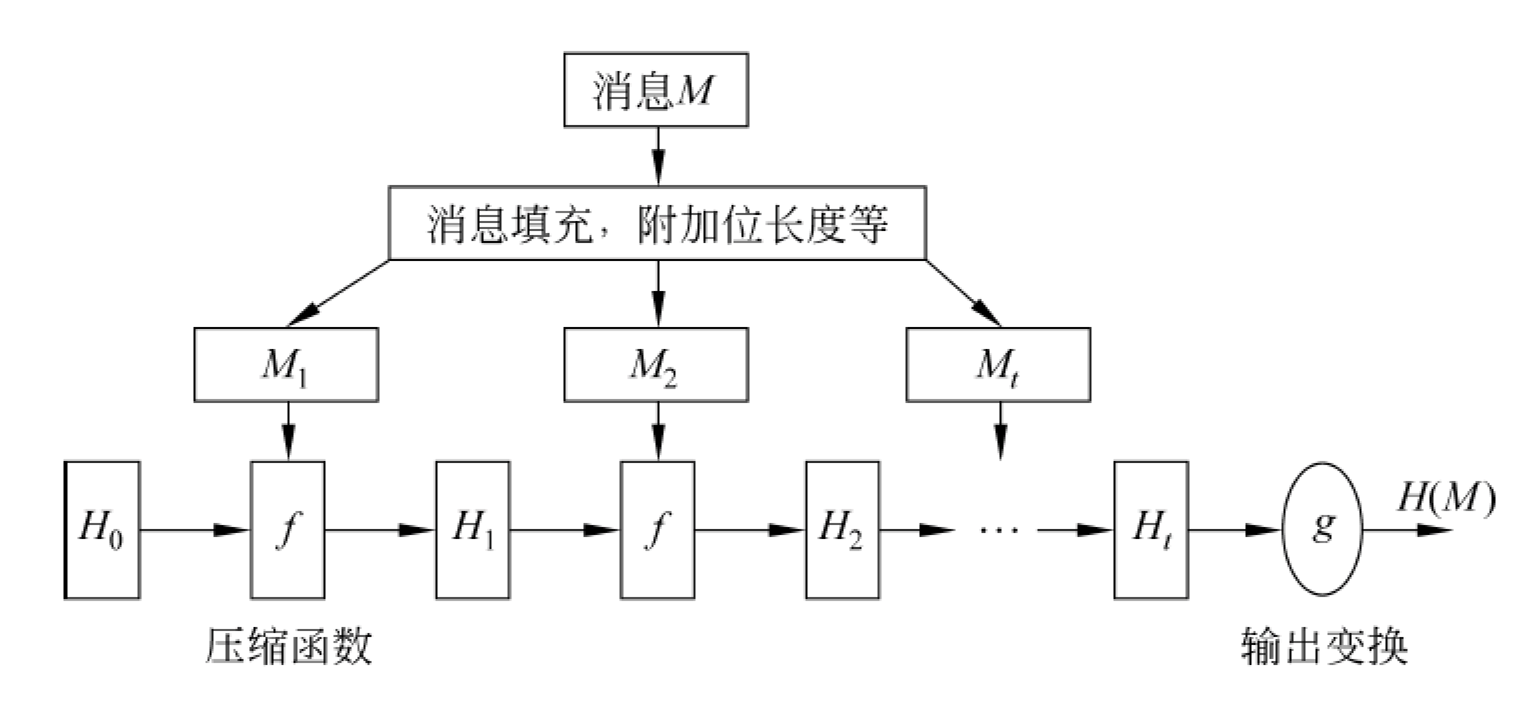


图 3-2 安全单向 Hash 函数的一般结构

$$M_i = f(H_{i-1}, M_i), \quad i = 1,2,\cdots,t \tag{3-1}$$

通常，H_0 为初始向量。

3.1.2 常见的单向 Hash 函数

常见的单向 Hash 函数包括 MD5、SHA-1、Tiger hash 和 CRC 等。

1. MD5

MD5(Message Digest 5)是 RSA 数据安全公司开发的一种单向 Hash 算法，MD5 可以用来把不同长度的数据块进行运算处理生成一个 128 位的数据块。

MD5 算法可简要地叙述为：MD5 以 512 位分组来处理输入的信息，且每一分组又被划分为 16 个 32 位的子分组，经过了一系列的处理后，算法的输出由四个 32 位分组组成，最终将这四个 32 位分组级联后生成一个 128 位 Hash 值。MD5 算法的总体框架图如图 3-3 所示。

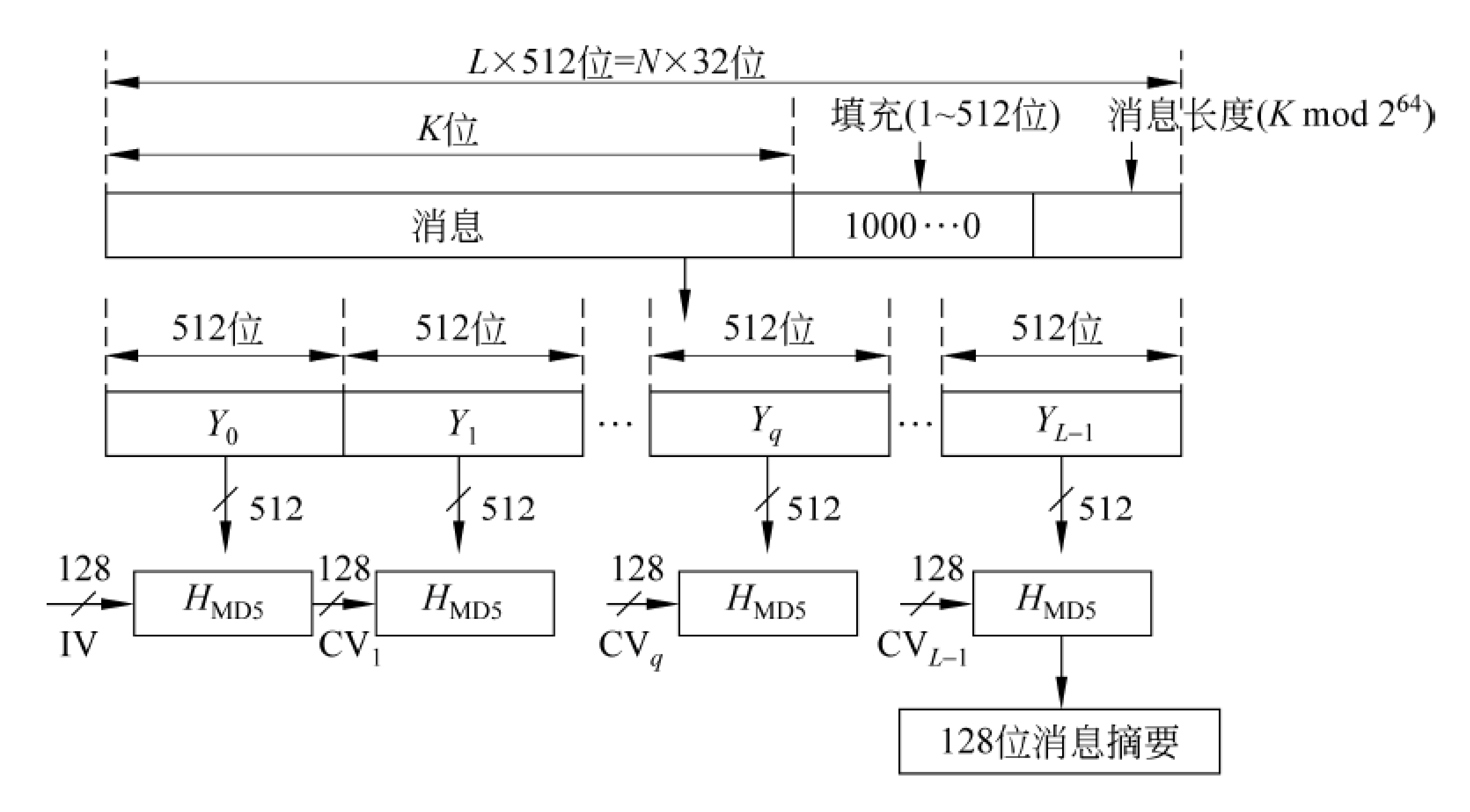


图 3-3 MD5 算法的总体框架图

在 MD5 算法中，首先需要对信息进行填充，使其位长度满足模 512 等于 448。因此，信息的位长度将被扩展至 $N\times512+448$，即 $N\times64+56$ 个字节(Bytes)，N 为一个非负整

数。填充的方法如下，在信息的后面填充一个 1 和无数个 0，直到满足上面的条件时才停止用 0 对信息的填充。然后，再在这个结果后面附加一个以 64 位二进制表示的填充前信息长度。经过这两步的处理，现在的信息字节长度 $=N\times512+448+64=(N+1)\times512$，即长度恰好是 512 的整数倍。这样做的原因是为满足后面处理中对信息长度的要求。对于单个信息分组，其处理过程如图 3-4 所示。

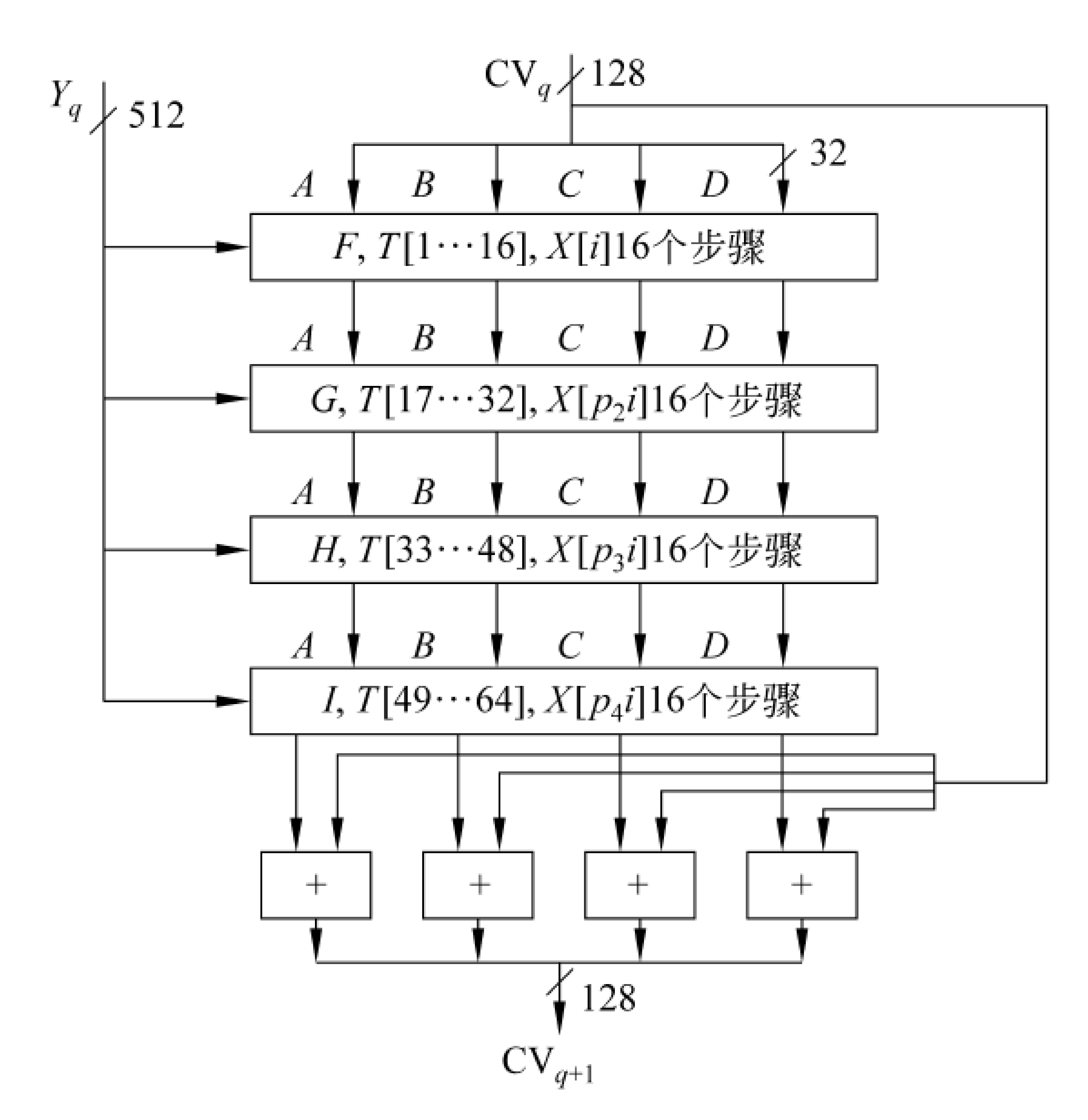


图 3-4　MD5 对单个 512 位分组的处理过程

MD5 中有 4 个 32 位被称作链接变量(chaining variable)的整数参数，它们分别为：A=0x01234567，B=0x89abcdef，C=0xfedcba98，D=0x76543210。

当设置好这四个链接变量后，就开始进入算法的四轮循环运算。循环的次数是信息中 512 位信息分组的数目。

主循环有四轮(MD4 只有三轮)，每轮循环都很相似。第一轮进行 16 次操作。每次操作对 A、B、C 和 D 中的其中三个作一次非线性函数运算，然后将所得结果加上第四个变量，文本的一个子分组和一个常数。再将所得结果向右环移一个不定的数，并加上 A、B、C 或 D 中之一。最后用该结果取代 A、B、C 或 D 中之一。

以下是每次操作中用到的四个非线性函数(每轮一个)。

$$F(X,Y,Z)=(X\&Y)\mid(\overline{X}\&Z)$$
$$G(X,Y,Z)=(X\&Z)\mid(Y\&\overline{Z})$$
$$H(X,Y,Z)=X\oplus Y\oplus Z$$
$$I(X,Y,Z)=Y\oplus(X\mid(\overline{Z}))$$

其中，& 是与，| 是或，－是非，⊕是异或。

这四个函数的说明：如果 X、Y 和 Z 的对应位是独立和均匀的，那么结果的每一位也应是独立和均匀的。F 是一个逐位运算的函数。即，如果 X，则 Y，否则 Z。函数 H 是逐位奇偶操作符。

每一轮都会使用到一个 64 元素表 $T[1\cdots64]$中的四分之一，$T[1\cdots64]$表是通过正弦函数构造得到的。T 中的第 i 个元素表示为 $T[i]$，它等于 $2^{32}\times\text{abs}(\sin(i))$的整数部分值，$i$ 的单位是弧度。

在 MD5 算法中，其核心是压缩函数 H_{MD5}。MD5 的压缩函数中有 4 次循环，每一次循环包含对缓冲区 ABCD 的 16 步操作，每一循环的形式为：

$$(a,b,c,d)=(d,b+((a+g(b,c,d)+X[k]+T[i])\ll s),b,c)$$

其中，a、b、c、d 对应着缓冲区 A、B、C、D 中的 4 个字；g 表示 F、G、H、I 中的某一个函数；$X[k]$表示当前 512 位数据块 Y_q 中的第 k 个 32 位；$\ll s$ 表示把 32 位循环左移 s 位；$+$是 mod 2^{32}。MD5 的基本操作如图 3-5 所示。

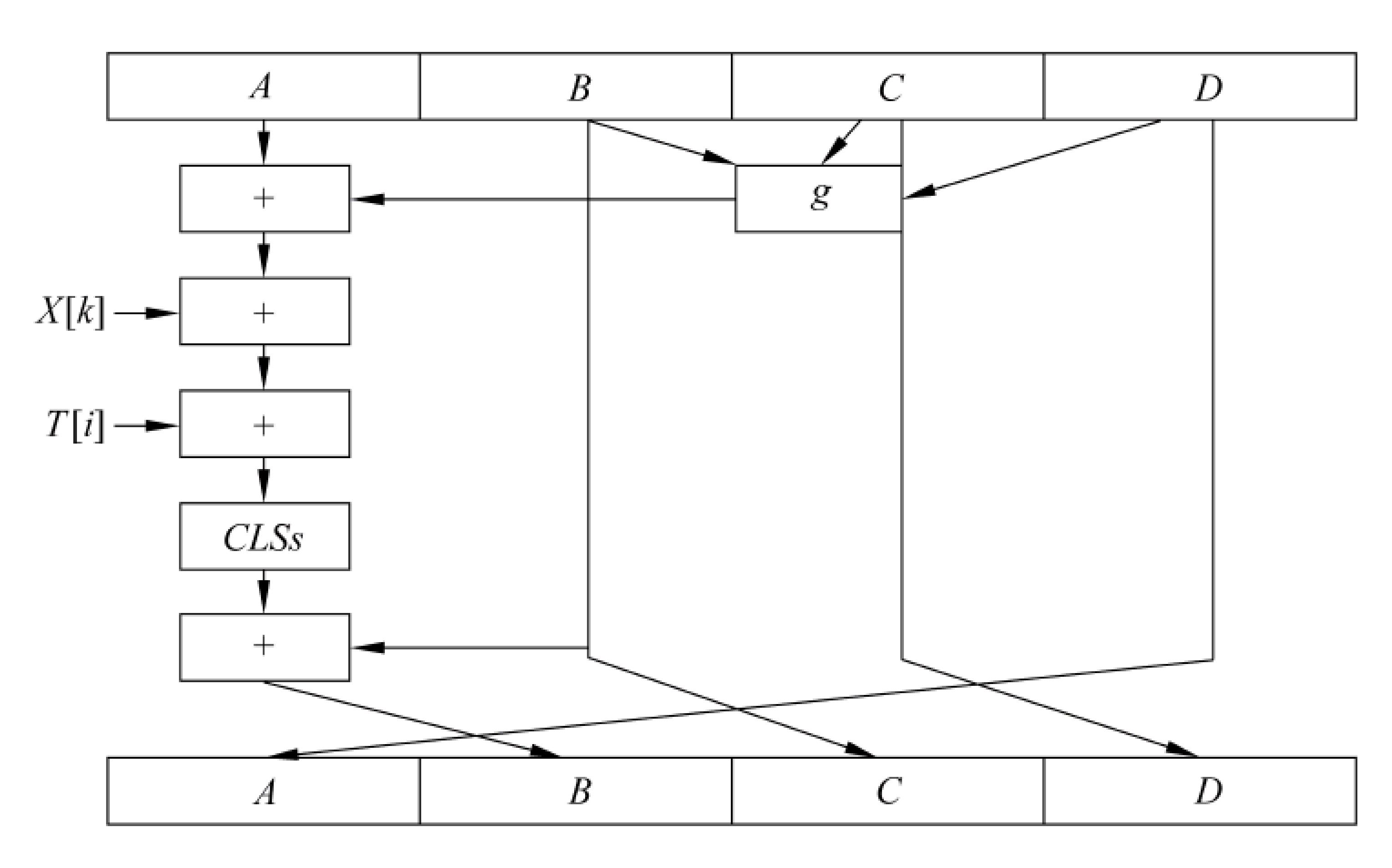


图 3-5　基本的 MD5 操作

2. SHA-1

安全 Hash 算法(SHA)是由美国 NIST 开发，作为联邦信息处理标准 FIPS PUB 180 于 1993 年发表，在 1995 年修订以后，称为 SHA-1(即 FIPS PUB 180-1 标准)。SHA-1 是基于 MD4 算法设计的。

SHA-1 主要适用于数字签名标准(Digital Signature Standard，DSS)里面定义的数字签名算法(Digital Signature Algorithm，DSA)。对于长度小于 2^{64}位的消息，SHA-1 会产生一个 160 位的消息摘要。当接收到消息的时候，这个消息摘要可以用来验证数据的完整性。在传输的过程中，数据很可能会发生变化，那么这时候就会产生不同的消息摘要。SHA-1 有这样的特性：不可以从消息摘要中复原信息；两个不同的消息不会产生同样的消息摘要。

SHA-1 算法的处理步骤如下：

① 添加填充位。SHA-1 算法对信息的填充和 MD5 采用的办法完全一样。

② 添加长度。一个 64 位的数据块，表示原始消息的长度。

③ 初始化消息摘要的缓冲区(即 IV 值)。消息缓冲区包括 160 位，用 5 个 32 位的寄存器(A，B，…，E)表示，用来存储中间和最终 Hash 函数的结果。初始化为(十六进制

表示)：

$$A = 0x67452301$$
$$B = 0xefcdab89$$
$$C = 0x98badcfe$$
$$D = 0x10325476$$
$$E = 0xc3d2e1f0$$

④ 以 512 位数据块作为单位来对消息进行处理。算法的核心是一个包含四个循环的模块，每个循环由 20 个处理步骤组成，其处理过程如图 3-6 所示。

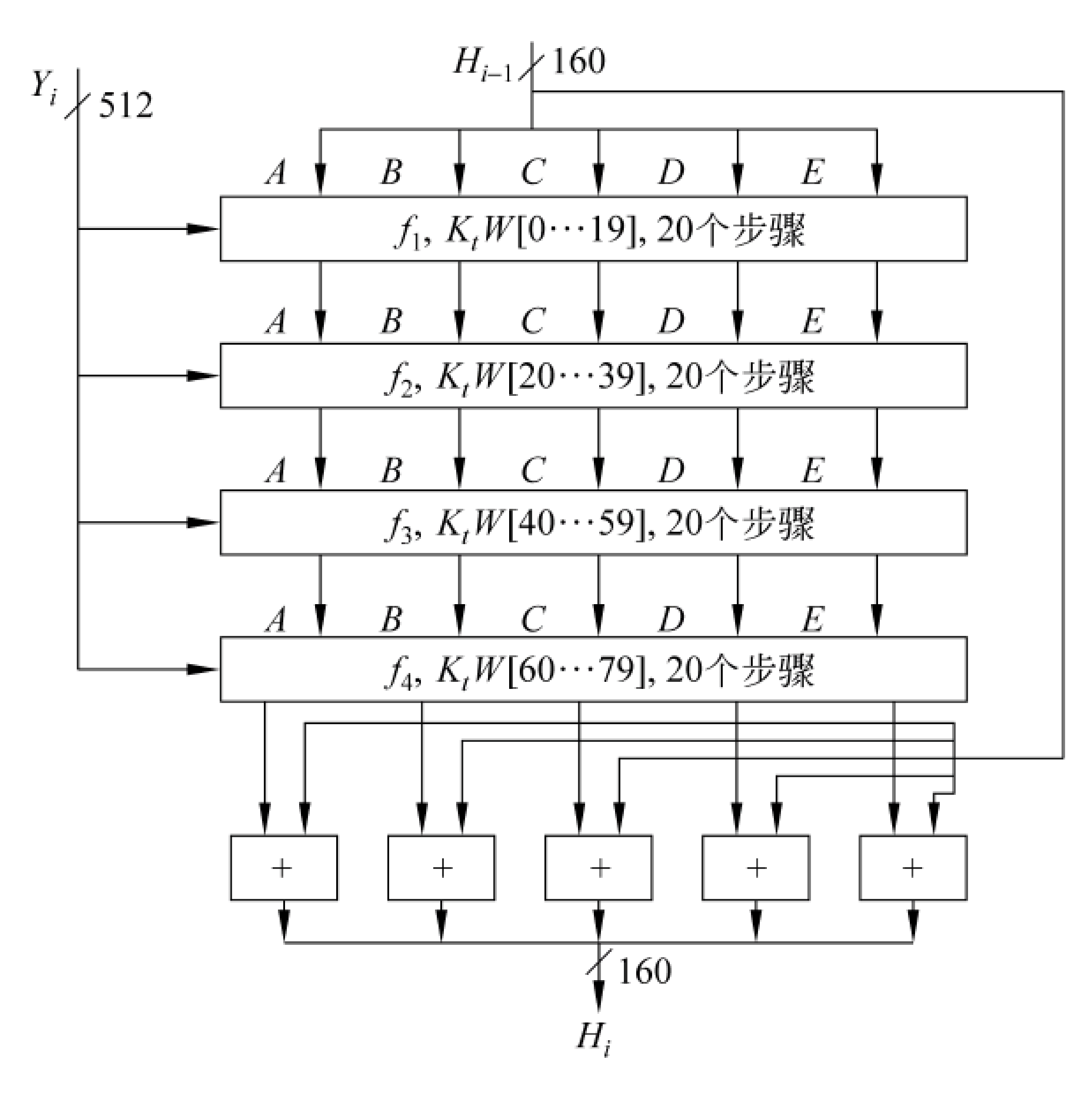


图 3-6 SHA-1 对单个 512 位分组的处理过程

图中的 f_1, f_2, f_3, f_4 为四个基本逻辑函数，它们的结构相似，每个循环使用不同的逻辑函数。逻辑函数的定义为：

$$f_1(t,B,C,D) = (B \wedge C) \mid (\bar{B} \wedge D) \quad (0 \leqslant t \leqslant 19)$$
$$f_2(t,B,C,D) = B \wedge C \wedge D \quad (20 \leqslant t \leqslant 39)$$
$$f_3(t,B,C,D) = (B \wedge C) \mid (B \wedge D) \mid (C \wedge D) \quad (40 \leqslant t \leqslant 59)$$
$$f_4(t,B,C,D) = B \wedge C \wedge D \quad (60 \leqslant t \leqslant 79)$$

K_t 为常量字，可用十六进制表示如下：

$$K_t = 0x5A827999 \quad (0 \leqslant t \leqslant 19)$$
$$K_t = 0x6ED9EBA1 \quad (20 \leqslant t \leqslant 39)$$
$$K_t = 0x8F1BBCDC \quad (40 \leqslant t \leqslant 59)$$
$$K_t = 0xCA62C1D6 \quad (60 \leqslant t \leqslant 79)$$

图 3-6 中的"+"表示 mod 2^{32}，Y_q 是指当前 512 位的消息分组。$W[j]$是由当前消息分组 Y_i 生成的一组字，总共 80 个。其生成规则为：$W[0]$～$W[15]$直接取自当前消息分组 Y_i

对应字的值,其他的定义如下:

$$W[t]=S^1(W[t-16]\oplus W[t-14]\oplus W[t-8]\oplus W[t-3])$$

其中,S^1 表示循环左移位一位操作。

SHA-1 的压缩函数如图 3-7 所示。SHA-1 的压缩函数可表示为:

$$(A,B,C,D,E)\leftarrow((E+f(t,B,C,D)+S^5(A)+W_t+K^t),A,S^{30}(B),C,D)$$

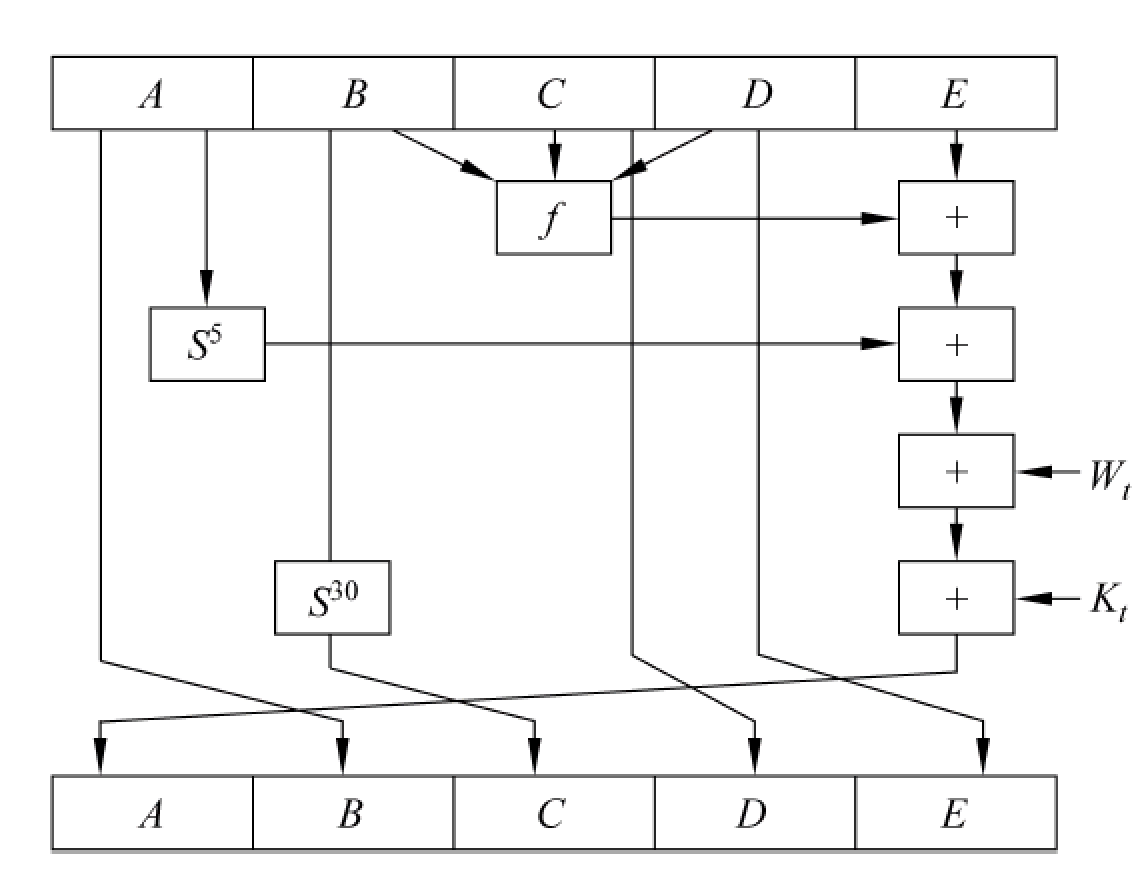


图 3-7 基本的 SHA-1 操作

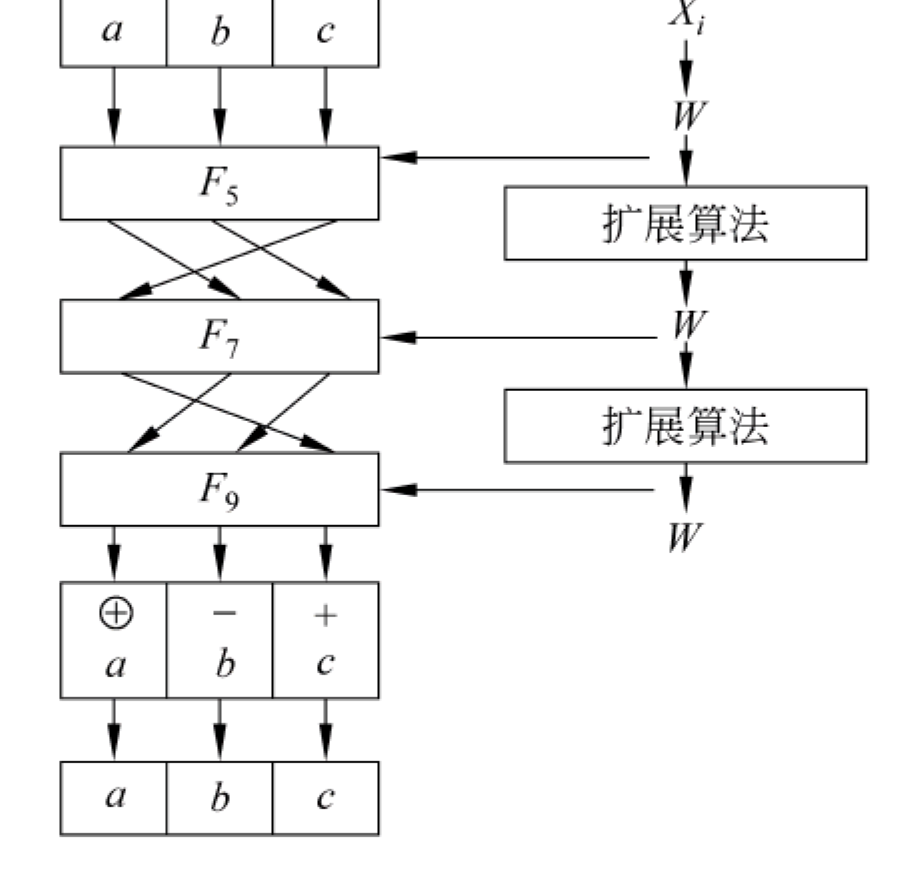


图 3-8 Tiger Hash 的外循环

3. Tiger Hash

MD5 和 SHA-1 的结构比较简单,都由一些随机的转换组成。Tiger Hash 函数是由 Ross Anderson 和 Eli Biham 提出的,结构比 MD5 和 SHA-1 更复杂。事实上 Tiger 的结构更接近于分组密码。

同 MD5 和 SHA-1 一样,Tiger 的输入被分成 512 位的分组,如有需要则将输入填充为 512 位的整数倍。与 MD5 和 SHA-1 不同的是 Tiger 的输出是 192 位。选择输出位数为 192 的设计目的是为了适应 64 位处理器,因为 192 正好是 64 位字。在 Tiger 中每轮的中间值也是 192 位。

从它使用的 4 个 S 盒就可以看出 Tiger 的设计受到分组密码的影响,每个 S 盒将 8 位映射成 64 位,Tiger 还应用了"密钥扩展"算法,这是因为没有密钥,实际上只是对输入分组进行扩展。

输入信息 X 被填充成 512 位的整数倍,然后写成:

$$X=(X_0,X_1,\cdots,X_{n-1})$$

这里每个 X_i 都是 512 位。Tiger 算法对每个 X_i 使用一个外循环,这里 $i=0,1,2,\cdots,n-1$,每轮的结构如图 3-8 所示。

a、b、c 都是 64 位,第一轮初始值(a,b,c)是:

$$a=0\text{x}0123456789\text{ABCDEF}$$

$$b=0\text{xFEDCBA}9876543210$$

$$c=0\text{xF}096\text{F}5\text{B}4\text{C}3\text{B}2\text{E}187$$

这里每轮的结果(a,b,c)作为下一轮的初始值。最后一轮的结果(a,b,c)就是192位Hash值。从这点上来看，Tiger与分组密码非常相似。

注意对于外循环F_5的输入是(a,b,c)。将F_5的输出标记为(a,b,c)，F_7的输出标记为(c,a,b)。图3-8中每个函数F_m由8个如图3-9所示的内循环构成。将512位输入W写成：

$$W = (W_0, W_1, \cdots, W_7)$$

这里每个W_i都是64位。图3-9中每条线都代表64位。

对于$i=0,1,2,\cdots,7$，$f_{m,i}$的输入值分别是：

$(a,b,c),(b,c,a),(c,a,b),(a,b,c),$

$(b,c,a),(c,a,b),(a,b,c),(b,c,a)$

$f_{m,i-1}$的输出分别标记为(a,b,c)，每个$f_{m,i}$依赖于a,b,c,W_i和m，这里的W_i是512位输入W的第i个64位子块。$f_{m,i}$的下标m是乘数。c写成：

$$c = (c_0, c_1, \cdots, c_7)$$

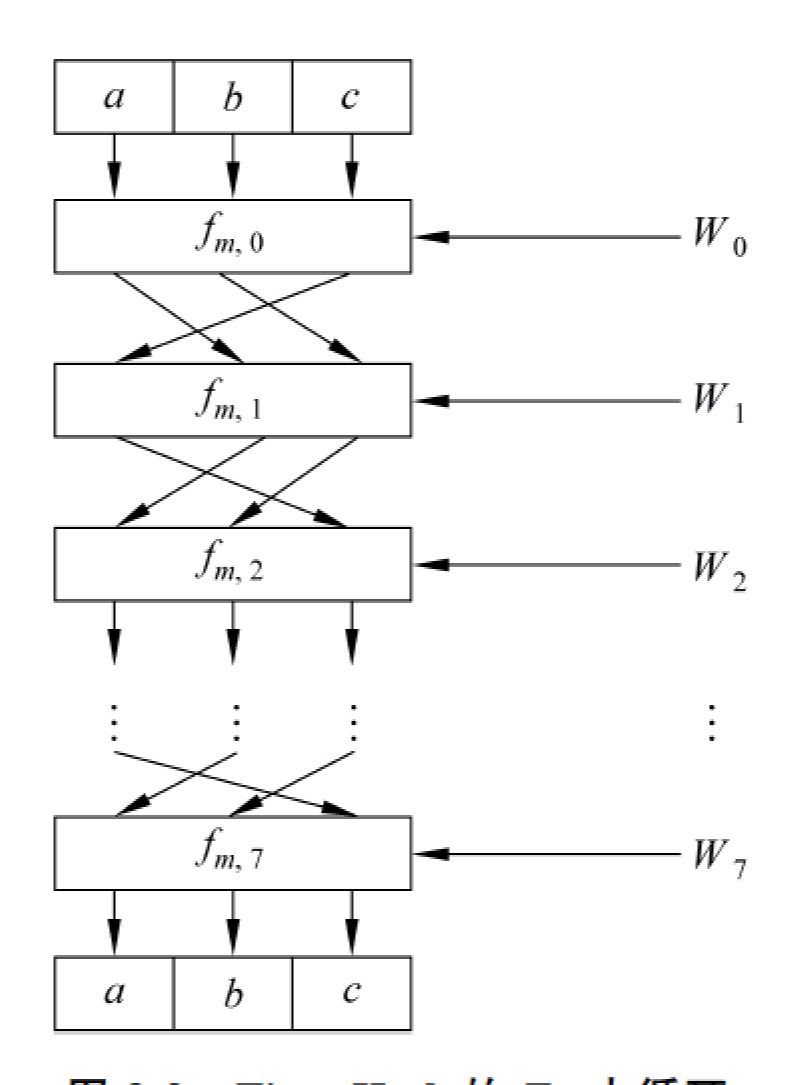


图3-9 Tiger Hash的F_m内循环

这里每个c_i都是单字节。$f_{m,i}$定义如下：

$$c = c \oplus \omega_i$$
$$a = a - (S_0[c_0] \oplus S_1[c_2] \oplus S_2[c_4] \oplus S_3[c_6])$$
$$b = b - (S_3[c_1] \oplus S_2[c_3] \oplus S_1[c_5] \oplus S_0[c_7])$$
$$b = b \cdot m$$

这里每个S_i都是8位映射到64位的S盒。由于这些S盒很大，这里就不给出了。

4. CRC

CRC(Cyclic Redundancy Check，循环冗余校验码)由于实现简单、检错能力强，被广泛使用在各种数据校验中。它占用系统资源少，用软件或硬件均能实现，是一种很好的进行数据传输差错检测的手段(CRC并不是严格意义上的Hash算法，但它的作用与Hash算法大致相同，所以归于此类)。

生成CRC码的基本原理：任意一个由二进制位串组成的代码都可以和一个系数仅为0和1取值的多项式一一对应。例如，代码1010111对应的多项式为$x^6+x^4+x^2+x+1$，而多项式为$x^5+x^3+x^2+x+1$对应的代码101111。

CRC码集选择的原则：若设码字长度为N，信息字段为K位，校验字段为R位($N=K+R$)，则对于CRC码集中的任一码字，存在且仅存在一个R次多项式$g(x)$，使得：

$$V(x) = A(x)g(x) = x^R m(x) + r(x)$$

其中，$m(x)$为K次原始的信息多项式，$r(x)$为$R-1$次校验多项式(即CRC校验和，由多项式$g(x)$对信息多项式$m(x)$做模2除得到)，$g(x)$称为生成多项式：

$$g(x) = g_0 + g_1 x + g_2 x^2 + \cdots + g_{R-1} x^{R-1} + g_R x^R$$

发送方通过指定的$g(x)$产生CRC码字，接收方则通过该$g(x)$来验证收到的CRC码字。

CRC 校验码生成方法：借助于多项式除法，其余数为校验字段。

例 3-1 信息字段代码为 1011001，相应地，$m(x)=x^6+x^4+x^3+1$，假设生成多项式为 $g(x)=x^4+x^3+1$；则对应 $g(x)$ 的代码为 11001。

$$x^4m(x) = x^{10} + x^8 + x^7 + x^4$$

对应的代码记为 10110010000。

采用多项式除法：得余数为：1010(即校验字段为：1010)。

发送方：发出的传输字段为：1 0 1 1 0 0 1 1010(信息字段 校验字段)。

接收方：使用相同的生成码进行校验：接收到的字段/生成码(二进制除法)。

如果能够除尽，则正确。下面给出余数(1010)的计算步骤。

除法没有数学上的含义，而是采用计算机的模二除法，即除数和被除数做异或运算。进行异或运算时除数和被除数最高位对齐，按位异或。

```
 10110010000
⊕11001
-----------
 01111010000
  1111010000
 ⊕11001
-----------
  0011110000
    11110000
   ⊕11001
-----------
    00111000
      111000
     ⊕11001
-----------
      001010
```

则四位 CRC 监督码就为：1010。

利用 CRC 进行检错的过程可简单描述为：在发送端根据要传送的 k 位二进制码序列，以一定的规则产生一个校验用的 r 位监督码(CRC 码)，附在原始信息后边，构成一个新的二进制码序列数共 $k+r$ 位，然后发送出去。在接收端，根据信息码和 CRC 码之间所遵循的规则进行检验，以确定传送中是否出错。这个规则，在差错控制理论中称为"生成多项式"。

3.1.3 单向 Hash 函数的攻击方法

对单向 Hash 函数的攻击可以分为普通攻击(generic attacks)和快捷攻击(short-cut attacks)。前者适用于所有的算法，其攻击的复杂度取决于 Hash 值的空间大小；而后者是通过利用某些特定算法的弱点，从而较普通攻击更容易对算法进行攻击。

1. 单向 Hash 函数的普通攻击

1) 随机(第二)原像攻击

随机(第二)原像攻击中，攻击者只是简单地随机选取一个输入从而获得所期望的输

出。如果单向 Hash 函数够“随机”的话,其攻击成功的概率为 $1/|R|$,这里 $|R|$ 表示 Hash 值的空间大小。可以通过多次选择输入、核对 Hash 值来提高攻击成功的机会。一般,大约需要进行 $r=O(|R|)$ 次这样的操作。当 $r=0.7|R|$ 时,攻击成功的概率约为 50%,当 $r=|R|$ 时,攻击成功的概率约为 63%。

2) 生日攻击

生日攻击来自于这样一个想法,在 23 个人中,其中两个人有相同生日的概率是 50%。由于在直觉上,大多数人会认为人数应该更多才会有达到这样的结果,故它被叫做生日悖论。对此的解释如下:由于一年有 365 天,假设在 r 个人中,每个人的生日都不相同,其概率为:

$$q=\frac{365\times364\times\cdots\times(365-r+1)}{365^r}=\prod_{i=0}^{r-1}\left(1-\frac{i}{365}\right)$$

故在该组人群中,至少有两个人有相同生日的概率为 $p=1-q$。当 $r=23$ 时,$p\approx0.507$。具有相同生日的概率随着该人群中人数的增加而增加。例如,当 $r=46$ 时,$p\approx0.948$。

在针对单向 Hash 函数的生日攻击中,攻击者通过任意选择 r 个不同的输入,期望能够得到至少两个输入具有相同的 Hash 值(也就是意味着发生了碰撞)。其概率的计算与前面介绍的方法类似,只是用 $|R|$ 代替 365(这必须假设单向 Hash 函数具有“随机”的特性)当 $r=O(\sqrt{|R|})$ 且 $|R|\to\infty$,发生碰撞的概率近似为:

$$p\approx1-\exp\left(-\frac{r^2}{2|R|}\right)$$

Flajolet P. 和 Odlyzko A. 计算出发生碰撞所需要的输入为:

$$r_{\text{col}}=\sqrt{\frac{\pi|R|}{2}}$$

3) Hash 值的安全输出长度分析

P. van Oorschot 和 M. Wiener 讨论了将生日攻击应用在 MD5 算法上所需要的开销。由于 MD5 的输出长度为 $n=128$ 比特,这意味着需要进行 2^{64} 次尝试才可能产生一次碰撞。在 1995 年的分析结果表明,用一台定制的价值 1 千万美元的机器可以在 21 天完成一次 MD5 的碰撞。根据电脑硬件发展的“摩尔定律”,每过 18 个月,硬件的计算能力将会增加一倍。显然,输出长度为 128 比特的单向 Hash 函数无法抵抗碰撞攻击。该分析结果同时也得到了其他研究人员的验证。1996 年,M. Blaze 指出,只有当生日攻击的复杂度达到 2^{75} 时才认为能够抵抗碰撞攻击。但是目前在实际应用中以及一些安全性要求较低的场合,一般认为 128 比特是安全的。

2. 对单向 Hash 函数的快捷攻击

攻击一个迭代单向 Hash 函数的难度与攻击其压缩函数的难度是相同的。因此,单向 Hash 函数的安全性可以由对其压缩函数的分析来衡量,单向 Hash 函数的设计者可以将他们的主要精力集中在压缩函数的设计上面。

链式攻击是对具有迭代特性的单向 Hash 函数的一种典型快捷攻击。该攻击主要是针对压缩函数,而不是整个单向函数。下面介绍几种不同的链式攻击方法。

1）定点攻击

给定 H、X，压缩函数 f 满足 $f(H,X)=H$，我们称压缩函数 f 含有定点。攻击者可以在计算的链值中出现 H 时，增加任意数目的值为 X 的信息块，就可以通过该性质进行第二原像攻击或者碰撞攻击。只有当初始链值为不固定（由攻击者选取 $IV=H$），或者恰巧能够找到产生定点的链值 H 的情况下，定点攻击才有可能发生。此外，该攻击只有在信息的补充信息中没有包含信息长度的情况下方能奏效，同时该攻击要求压缩函数的迭代次数要多于1次。

2）变更块攻击

假定攻击者想寻找一个给定的包含 t 块的输入 X 的第二原像，也就是说是要寻找 X' 满足 $H(X')=H(X)$。在变更块攻击中，攻击者选择输入的某一个块 X_i，然后用另外一个块 X_i' 并使得 $f(H_i,X_i')=f(H_i,X_i)$ 成立。如果 X' 中的其他的块和 X 的相应块相同，则 X' 和 X 的 Hash 值相等，从而完成了对单向 Hash 函数的第二原像攻击。

3）Meet-in-the-middle 攻击

Meet-in-the-middle（中途相遇）攻击是生日攻击的一个变种。在生日攻击中，每一次尝试是比较最终的 Hash 值是否相同，而在该攻击中，只是比较中间的链值是否相同。在应用时，Meet-in-the-middle 攻击可以让攻击者进行原像攻击，这一点无法在生日攻击中实现。

在 MD5 算法被以王小云为代表的中国专家攻破后，世界密码学界仍然认为 SHA-1 算法是安全的。然而，2005 年 2 月，王小云就宣布攻破 SHA-1 算法的消息。因为 SHA-1 在美国等国家有更加广泛的应用，密码被破的消息一出，在国际上的反响可谓石破天惊。换句话说，王小云的研究成果表明了 Hash 值从理论上讲是可以伪造的，必须及时添加限制条件，或者重新选用更为安全的密码标准，以保证电子商务的安全。

3.2 消息认证码

3.2.1 基本概念

消息认证码（Message Authentication Code，MAC）也叫密码校验和（cryptographic checksum）是鉴别函数的一种。

消息认证码实现鉴别的原理是：用公开函数和密钥产生一个固定长度的值作为认证标识，用这个标识鉴别消息的完整性。使用一个密钥生成一个固定大小的小数据块，即 MAC，并将其加入到消息中，然后传输。接收方利用与发送方共享的密钥进行鉴别认证等。

消息认证实际上是对消息本身产生一个冗余的信息——MAC（消息认证码），消息认证码是利用密钥对要认证的消息产生新的数据块并对数据块加密生成的。它对于要保护的信息来说是唯一的且一一对应的。因此可以有效地保护消息的完整性，以及实现发送方消息的不可抵赖和不能伪造。消息认证码的安全性取决于两点：一是所采用的加密算法；二是待加密数据块的生成方法。

消息认证不支持可逆性,是多对一的函数,其定义域由任意长的消息组成,而值域则是由远小于消息长度的比特值构成,从理论上说,一定存在不同的消息产生相同的冗余数据块。因此必须要找到一种足够单向和强碰撞自由性的方法对消息认证才是安全的。

(1) 利用校验码加密的方式构造认证码,它可以实现数据的完整性,它对消息不可抵赖、不可伪造性的认证性能取决于加密的函数。因此这种方法的安全性取决于校验码的长度和加密的方法。但是由于它是针对局部变量的校验,比如针对一行或者一列,它的抗碰撞性能不是很好,即有可能产生消息被改动,而认证码仍然没有变动的情况。

(2) 对于用单向 Hash 函数构造认证码的方式来说,安全性是基于该函数的抗强碰撞性的,即攻击的主要目标是找到一对或更多对碰撞消息,该消息生成摘要是相同的。在目前已有的攻击方案中,一般的方法是基于穷举的,可攻击任何类型的 Hash 方案,例如生日攻击方法。另一些是特殊的方法,只能用于攻击某些特殊类型的 Hash 方案,例如适用于攻击具有分组链结构的 Hash 方案的中间相遇攻击,适用于攻击基于模算术的 Hash 函数的修正分组攻击。因此摘要的长度是一个关键的因素。

3.2.2 常见的消息认证码算法

MAC 实质上是一个将双方共享的密钥 k 和消息 m 作为输入的函数,如果将函数值记为 $\mathrm{MAC}_k(m)$,这个函数值就是一个认证标记,这里用 δ 表示。攻击者发起攻击的时候,所能得到的是消息和标记的序列对 $(m_1,\delta_1),(m_2,\delta_2),\cdots,(m_q,\delta_q)$(其中 $\delta_i=\mathrm{MAC}_k(m_i)$)。如果攻击者可以找到一个消息 m,m 不在 $m_1,m_2,\cdots,m_q$ 之中,并且能够得到正确的认证标记 $\delta=\mathrm{MAC}_k(m)$ 就说明攻击成功了。攻击者成功的概率就是其攻破 MAC 的概率。

MAC 的构造方法有很多,一种是基于带密钥的 Hash 函数的;另一种是基于流密码的,这种 MAC 不多也不流行,这里不作讨论。另外,还有一种称为 Carter-Wegman MACs(首先使用一个泛 Hash 函数(universal Hash)将长消息散列成较短的字串,然后加密这个字串得到标记),这种 MAC 基于的思想和前两类没有明确的界限,有些使用前两种方法构造的 MAC 也可看成是 Carter-Wegman MACs。这里主要讨论基于带密钥的 Hash 函数。

Hash 函数可以把较长的消息变换为较短的消息摘要,并且具有抗碰撞性好的性质。为了保证消息的完整性,必须加入秘密信息——密钥,在加入了密钥之后,Hash 函数就称为带密钥的 Hash 函数。但是单独一个带密钥的 Hash 函数是不能直接作为 MAC 使用的,它必须经过特殊的构造,在具备了较好的性质后才可以用作消息认证码。在许多网络协议中,有很多使用这种方法构造的消息认证码,但是这些方法都是基于特殊技巧,很难进行安全性的分析和证明。这里介绍的基于带密钥的 Hash 函数的 MAC 可被证明是安全的。

基于带密钥的 Hash 函数的构造方法最早是由 M. Bellare 等人提出。它要求所使用的 Hash 函数具有迭代结构(如 MD5、SHA-1 等)。所谓迭代结构就是反复使用压缩函数 f 将长消息映射为短消息。这个压缩函数 f 具有两个输入,一个是长度为 l 的链变量,一

个是长度为 b 的数据块,表示为 $f_k=f(k,x)$,其中,k 的长度为 l,x 的长度为 b。在 MD5 中,$b=512$,$l=128$。

假定消息 $x=(x_1,x_2,\cdots,x_n)$,其中,x_i 是长度为 b 的块,$i=1,2,\cdots,n$,n 是总块数。x_{n+1} 也是长度为 b 的二进制串,其中包含了 x 最后不足 b 的部分和整个消息长度的二进制表示以及一些填充位。那么使用压缩函数 f 构造的 Hash 函数 $F(x)$ 如图 3-10 所示。

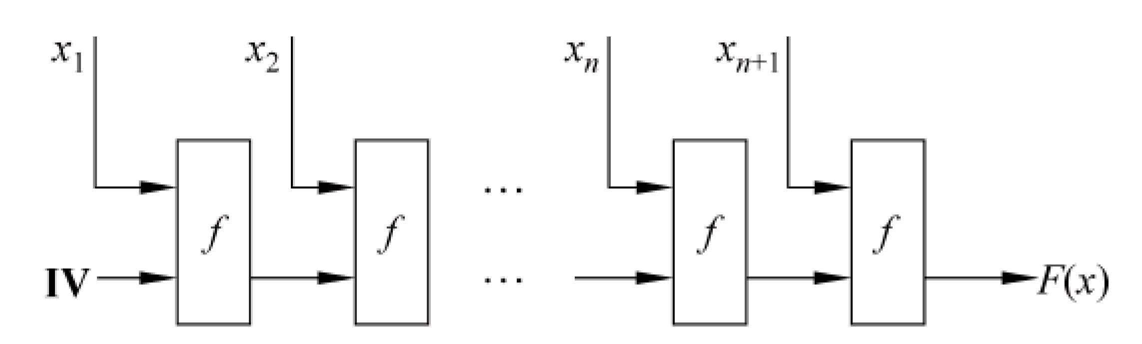


图 3-10 具有迭代结构的单向 Hash 函数

IV 代表初始向量,其长度为 l。如果使用密钥 k 作为初始向量,Hash 函数就成了带密钥的 Hash 函数。

如果让 F 代表初始向量固定为 **IV** 且具有迭代结构的 Hash 函数,那么 HMAC(x,k) 的构造方法如下:

```
function HMAC(x,k)
t←0,k̄←pad(x), x←(k̄⊕Ipad)||x
t←F(x), δ←F((k̄⊕Opad)||t)
return δ
```

其中,x 是任意长度的输入,k 是长为 l 的密钥,$\bar{k}$ 是密钥 k 填充到 b 位长之后的值。Opad 和 Ipad 是两个固定的长度为 b 的串。Opad 是 0x36 的重复直到 b 位长,Ipad 是 0x5c 的重复直到 b 位长。符号“‖”表示将两个二进制串串联起来。HMAC 已经取代了 RFC 1828 成为 IPSec 协议中的认证算法。

这种构造方法具备很多优点,和同类型的 MAC 算法相比,它给出了安全性证明将 MAC 的安全性归结到所使用 Hash 函数上。在软件实现上,它要比使用分组密码构造的 MAC 快,而且它的效率特别高。从它的构造上可以看出,它以一种非常简单的方式使用带密钥的 Hash 函数,同底层的 Hash 函数相比,性能并没有降低多少。另外两个值得称道的优点是免费和黑盒。免费是指使用 Hash 函数不受法律限制,可以免费使用。黑盒是指可以将底层的 Hash 函数看成一个模块,可根据需要方便地进行更换。

同时,这种构造方法还存在着不足,安全性依赖于底层的 Hash 函数,而所使用的 Hash 函数有些是没有安全性证明的,所以不能保证这种方法的安全性。其次,由于压缩函数是串行的,该构造方法不支持并行。

UMAC 是由 Black 等人使用 Carter-Wegman 提出的思想构造的一种算法。在这里将其划分为使用带密钥的 Hash 函数构造的 MAC,是因为该算法同样使用了带密钥的 Hash 函数。该算法首先使用 NH HASH 函数并充分利用计算机的计算特点将源消息变换为原来消息长度的 $2/n$(其中 n 为子密钥的个数),然后再对所产生的消息使用 HMAC(SHA)进行 Hash 变换。这种算法被认为是下一代的 MAC,其优点是速度很快,

缺点是处理变化的长度需要进行特殊的处理。

3.2.3 分组加密与消息认证码

基于分组密码设计的这一类 MAC 主要有：CBC-MAC、XOR-MAC、PMAC、XECB-MAC、OCB 和 EMAC(加密的 CBC-MAC)等。

1. CBC-MAC

CBC-MAC 其实就是对消息使用 CBC 模式进行加密，取密文的最后一块作为认证标记。具体的构造方法如下：

```
function  CBC-MAC(x,k)
y0←0^l, pad(x)
partition x into x1,…,xn
for i←1 to n
     yi←Fk(xi⊕yi-1)
     δ←yn
return δ
```

其中，x 为消息，k 为密钥，F 为某种分组密码算法，δ 就是所产生的标记。

这种方法出现得较早，是一种经典的构造方法，其构造方法简单，底层加密算法具备黑盒的性质，可以方便地进行替换。后来的很多 MAC 算法都是对它的改进。但是 CBC-MAC 仅适用于对相同长度的消息进行认证，在消息长度变化的情况下是不安全的。这些在文献[5]中都已经给出了证明，另外，它的构造方法决定了该算法不支持并行计算。

为了克服 CBC-MAC 的上述弱点对 CBC-MAC 进行了改进。Bellare 给出了三种方法，分别是 Input-length key separation、Length-prepending 和 Encrypt last block。其中最有效的方法是最后一种，也就是 EMAC。它是由 RIPE Project 在 1993 年提出的，接着被列入 ISO 标准中。它的具体构造是：$\mathrm{EMAC}_{E_{k_1},E_{k_2}}(x)=E_{k_2}(\mathrm{CBC}_{E_{k_1}}(x))$，其中 k_1，k_2 是密钥空间中两个不同的密钥。通信双方使用一个安全的密钥 K 产生这两个密钥 $k_1=E_k(0^l)$，$k_2=E_k(1^l)$，并且在证明这个 MAC 的安全性的时候认为 E_{k_1}，E_{k_2} 是两个独立随机选择函数。随后对 EMAC 进行了改进，得到了三种新的 MAC 算法：ECBC、FCBC 和 XCBC。

2. XOR-MAC

XOR-MAC 有两种方式：无状态(XMACR)和有状态(XMACC)。这种算法在计算过程中引入索引值使得分组密码每次加密的明文各不相同，最后再将所有的密文异或。具体的构造方法描述如下。

假定 $|x|$ 代表消息 x 的长度(即包含多少位)，并且它是 32 的倍数。$x=(x_1,x_2,\cdots,x_n)$，其中，$|x_i|=32$，$i=1,2,\cdots,n$。假定 n 小于 2^{31}。$<i>$ 是数字 i 长度为 b 的二进制表示，i 代表块的索引号。发送者保留一个长度为 63 位的记数 r，在 XMACC 模式下它的初始值为 0，每次增加 1。在 XMACR 模式下，r 是随机选取的一个长度为 63 位的串。它们

的构造如下：

```
function XMACR(x,k)
pad(x) r←{0,1}^63
y0=Fk(0||r)
partition x into x1,…,xn
for  i=1 to n
   yi=yi-1⊕Fk(1||<i>||xi)
return (r,yn)
```

XMACC 的构造如下：

```
function XMACC(x,k)
pad(x) ctr←ctr+1
y0=Fk(0||ctr)
partition x into x1,…,xn
for  i=1 to n
   yi=yi-1⊕Fk(1||<i>||xi)
return (ctr,yn)
```

由于 XOR-MAC 使用异或来生成认证码，这就为其带来了并行性、增量式、乱序验证(在验证的时候不需要按照顺序进行)等优点。在计算速度方面，基于 DES 的 XOR-MAC 在硬件实现效率上高于 CBC-MAC；在软件实现上使用 MD5 实现 XOR-MAC 效率较高。在安全性方面，它的安全性要高于 CBC-MAC。对攻击者来说，在理想情况下攻击 XOR-MAC 成功的概率要比攻击 CBC-MAC 成功的概率低，并且这个概率跟消息长度没有关系。该算法的主要缺点是在算法中引入了索引信息，导致了消息的扩展，导致了加密次数的增加，降低了运算速度。

3. PMAC

PMAC 可以看成是对 XOR-MAC 的改进，它也采用了异或来得到 MAC。它具有可并行、支持消息的添加、截短和替换等优点。它与 XOR-MAC 相比有两点不同：第一是所加密的内容不同，XOR-MAC 所加密的内容是消息连接上一个索引信息，而 PMAC 使用的是消息和不同的串进行异或之后的值。第二是对最后一块消息的处理不同。XOR-MAC 并不对最后一块消息进行特殊处理，而 PMAC 并不直接加密最后一块，而是先填充并和前面块的加密结果进行异或，然后再分情况进行处理，最后再加密一次。该算法使用了灰码(Gray Code)和有限域 GF(2)上的乘法运算。

PMAC 的产生方式是在线的，也就是说在计算 MAC 的时候不需要事先知道消息的长度。另外，PMAC 是确定性的，它不需要一个随机数或保留一个计数。虽然 PMAC 具备这么多的优点，但是它的速度比 CBC-MAC 要慢，且该算法受专利保护，不能免费使用。

4. XECB-MAC

XECB-MAC 也可看成是 XOR-MAC 的一种改进，它仍然采用异或的方法得到

MAC,因此它同时具有 XOR-MAC 的优点,如支持并行计算、增量式操作、乱序验证等。和 XOR-MAC 不同的是它没有使用消息的有效位来记录消息的位置,这样就减少了加密的次数,因此它的速度要高于 XOR-MAC,但低于 CBC-MAC。而且它的安全性没有 XOR-MAC 的高。由于在许多需要加密的情况下也同时需要对消息进行认证,而简单地将加密算法和认证算法结合起来的方法并不能保证其安全性,所以就出现了同时提供加密和认证的模式,这种模式有 XCBC 和 OCB 等。XCBC 对消息同时提供加密和认证,它也分为无状态和有状态两种。该方法支持实时的消息认证,所谓实时是指当加密完成时,认证标记就产生了。此外,该方法还具有支持并行计算等优点。该方法的不足之处在于使用了两个密钥,这给密钥的存储和分发带来了困难;而且所提供的完整性服务仅仅是对加密的一种补充,如果作为 MAC 单独使用,则会造成计算资源的浪费。

5. OCB

OCB 是在综合了 PMAC 和 XCBC-MAC 的构造方法的基础上提出来的,它同时提供了加密和认证。从构造方法上可以看出它与 PMAC 有一定的渊源,区别在于 OCB 同时提供加密和认证,而 PMAC 仅仅提供认证。OCB 的优点包括它能处理任意长度的消息,运算速度快,并且支持并行处理。该模式在同时需要保证消息的私密性和完整性的情况下适用,例如可以用在 SSL 和 SSH 协议中以取代当前使用的组合算法。OCB 的缺点在于算法复杂且不能免费使用。

3.3 数字签名技术

3.3.1 基本概念

数字签名(digital signature,又称公钥数字签名、电子签章)是一种类似写在纸上的普通的物理签名,但是它使用了公钥加密领域的技术实现,是一种用于鉴别数字信息的方法。一套数字签名通常定义为两种互补的运算:一种用于签名,另一种用于验证。数字签名不是指将签字者的签名扫描成数字图像,或者用触摸板获取的签名,更不是签字者的落款。

经过数字签名的文件的完整性是很容易验证的(不需要骑缝章、骑缝签名,也不需要笔迹专家),而且数字签名具有不可抵赖性(不需要笔迹专家来验证)。

简单地说,所谓数字签名就是附加在数据单元上的一些数据,或是对数据单元所作的密码变换。这种数据或变换允许数据单元的接收者用以确认数据单元的来源和数据单元的完整性并保护数据,防止被人(例如接收者)伪造。它是对电子形式的消息进行签名的一种方法,一个签名消息能在一个通信网络中传输。显然,数字签名的应用涉及法律问题,美国联邦政府基于有限域上的离散对数问题制定了自己的数字签名标准(DSS)。

数字签名技术是不对称加密算法的典型应用。数字签名的应用过程是,数据源发送

方使用自己的私钥对数据校验和或其他与数据内容有关的变量进行加密处理，完成对数据的合法“签名”，数据接收方则利用对方的公钥来解读收到的“数字签名”，并将解读结果用于对数据完整性的检验，以确认签名的合法性。数字签名技术是一种在网络系统虚拟环境中确认身份的重要技术，完全可以代替现实过程中的“亲笔签字”，在技术和法律上有保证。在数字签名应用中，发送者的公钥可以很方便地得到，但其私钥则需要严格保密。

3.3.2 常用的数字签名体制

基于公钥密码体制和私钥密码体制都可以获得数字签名，目前主要是基于公钥密码体制的数字签名。包括普通数字签名和特殊数字签名。普通数字签名算法有 RSA、DSS、ElGamal、Fiat-Shamir、Guillou-Quisquarter、Schnorr、Ong-Schnorr-Shamir 数字签名算法、DES/DSA、椭圆曲线数字签名算法和有限自动机数字签名算法等。特殊数字签名有盲签名、代理签名、群签名、不可否认签名、公平盲签名、门限签名、具有消息恢复功能的签名等，它与具体应用环境密切相关。

RSA 算法在第 2 章中已经介绍，在此就不再赘述。下面详细介绍 DSS 和 DSA 算法。DSS 最初提出于 1991 年，1993 年根据公众对安全性的反馈意见进行了一些修改，2000 年发布了该标准的扩充版，即 FIP 186-2。其最新版本还包括基于 RSA 和椭圆曲线密码的数字签名算法。

DSS 使用的是只提供数字签名的算法，与 RSA 不同，DSS 是一种公钥方法，但不能用于加密或密钥分配。图 3-11 对用 DSS 数字签名和 RSA 产生的数字签名这两种方法进行了对比，在 RSA 方法中，Hash 函数的输入是要签名的消息，输出是定长的 Hash 码，用发送方的私钥对该 Hash 码加密形成签名，然后发送消息及签名，接收方用发送方的公钥对签名进行解密，如果计算出的 Hash 码与解密出的结果相同，则认为签名是有效的。因为只有发送方拥有私钥，因此只有发送方能够产生有效的签名。

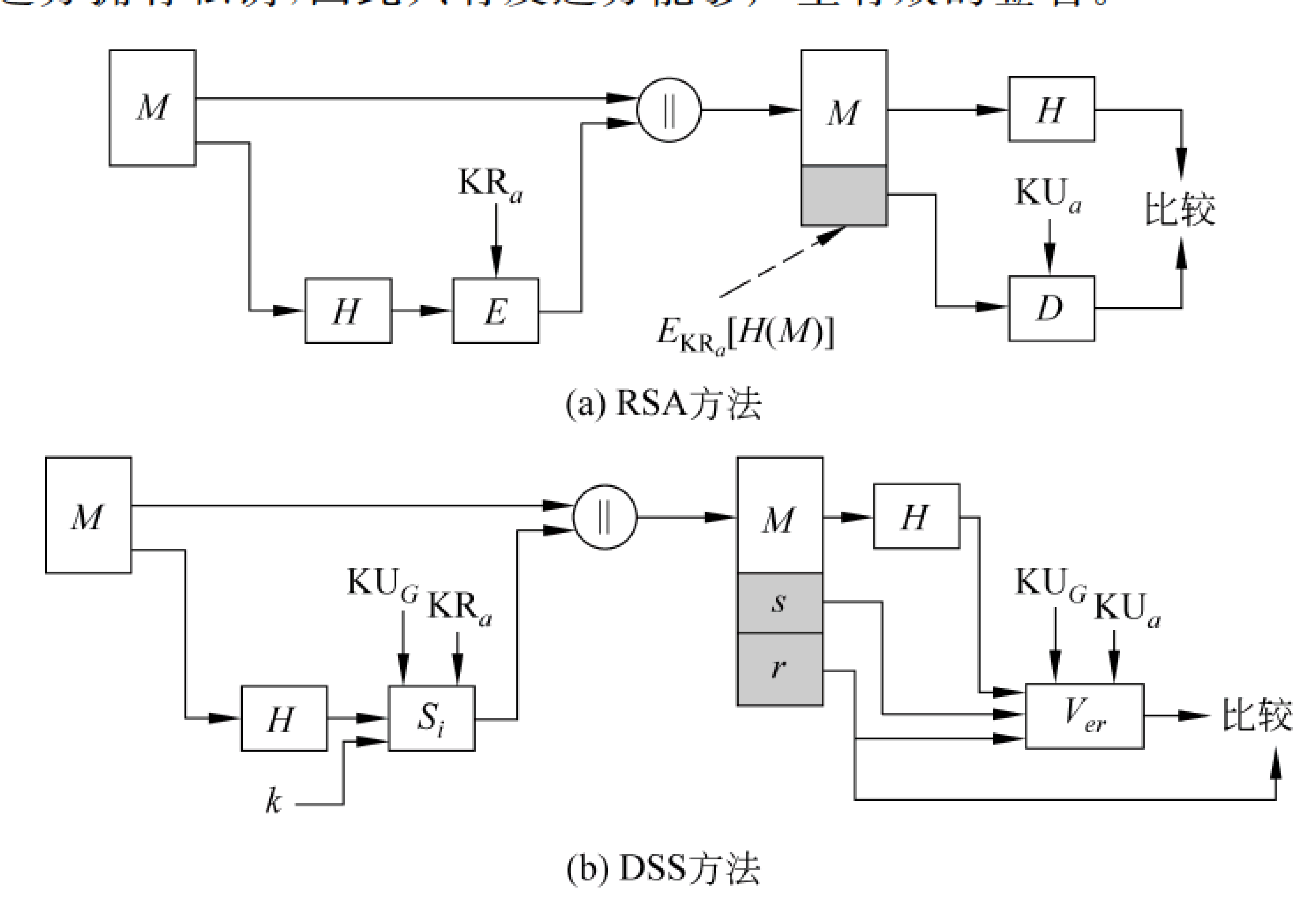


图 3-11 两种数字签名方法

DSS 方法也是用 Hash 函数，它产生的 Hash 值和为此次签名而产生的随机数 k 一起作为签名函数的输入，签名函数依赖于发送方的私钥(KR_A)和一组参数，这些参数为通信多方所共有，可以认为这组参数构成了全局公钥(KU_A)。签名由两部分组成，分别记为 s 和 r。

接收方对接收到的消息产生 Hash 码，这个 Hash 码和签名一起作为验证函数的输入，验证函数依赖于全局公钥和发送方公钥，若验证函数的输出等于签名中的 r 成分，则签名是有效的。签名函数保证只有拥有私钥的发送方才能产生有效签名。

DSA 建立在求离散对数的困难性以及 ElGamal 和 Schnorr 最初提出的方法之上。图 3-12 归纳总结了 DSA 算法，其中有三个公开参数为一组用户所共用。选择一个 160 位的素数 q，然后选择一个长度在 512～1024 之间，且满足 q 能整除($p-1$)的素数 p，最后选择形为 $h^{(p-1)/q} \bmod p$ 的 g，其中 h 是 1 到 $p-1$ 之间的整数，且 g 大于 1。

全局公钥组成

P为素数，其中$2^{L-1}<P<2^L$, $2^9 \leqslant P \leqslant 2^{10}$，且$L$ 是64 的倍数

q是$(p-1)$的素因子，其中 $2^{159} \leqslant q \leqslant 2^{160}$, 即位长为160 位

$g=h^{(p-1)/q} \bmod p$, 其中h满足$1<h<(p-1)$

并且$g=h^{(p-1)/q} \bmod p>1$

用户的私钥

x为随机或伪随机整数且$0<x<q$

用户的公钥

$y=g^x \bmod p$

与用户每条消息相关的秘密值

k为随机或伪随机整数且$0<k<q$

签名

$r=(g^k \bmod p) \bmod q$

$s=[k^{-1}(H(M)+xr)] \bmod q$

签名$=(r, s)$

验证

$w=(s')^{-1} \bmod q$

$u_1=[H(M')w] \bmod q$

$u_2=(r'w) \bmod q$

$v=[(g^{u_1} y^{u_2}) \bmod p] \bmod q$

检验：$v=r'$

M：要签名的消息

$H(M)$使用SHA−1计算的M的Hash值

M', r', s'接收到的M, r, s

图 3-12 数字签名算法(DSA)

选定这些参数后，每个用户选择私钥并产生公钥。私钥 x 必须是随机或伪随机选择的介于 1 到 $q-1$ 之间的数，可通过 $y=g^x \bmod p$ 计算得到公钥。由给定的 x 计算 y 比较简单，而由给定的 y 计算 x 则在计算上不可行，这就是求 y 的以 g 为底的模 p 的离散对数。

要进行签名，用户需计算两个量 r 和 s，r 和 s 是公钥(p,q,g)、用户私钥(x)、消息的 Hash 码 $H(M)$和附加整数 k 的函数，其中，k 是随机或伪随机产生的，且 k 对每次签名是唯一的。

图 3-13 更加详细地描述了上述签名和验证函数。该算法的特点为：接收端的验证依赖于 r，但是 r 却根本不依赖于消息，它是 k 和全局公钥的函数。k 模 p 的乘法逆元传给函数 f_1，f_1 的输入还包含消息 Hash 值和用户私钥。函数的这种结构使接收方可利用其收到的消息和签名、它的公钥以及全局密钥来恢复 r。

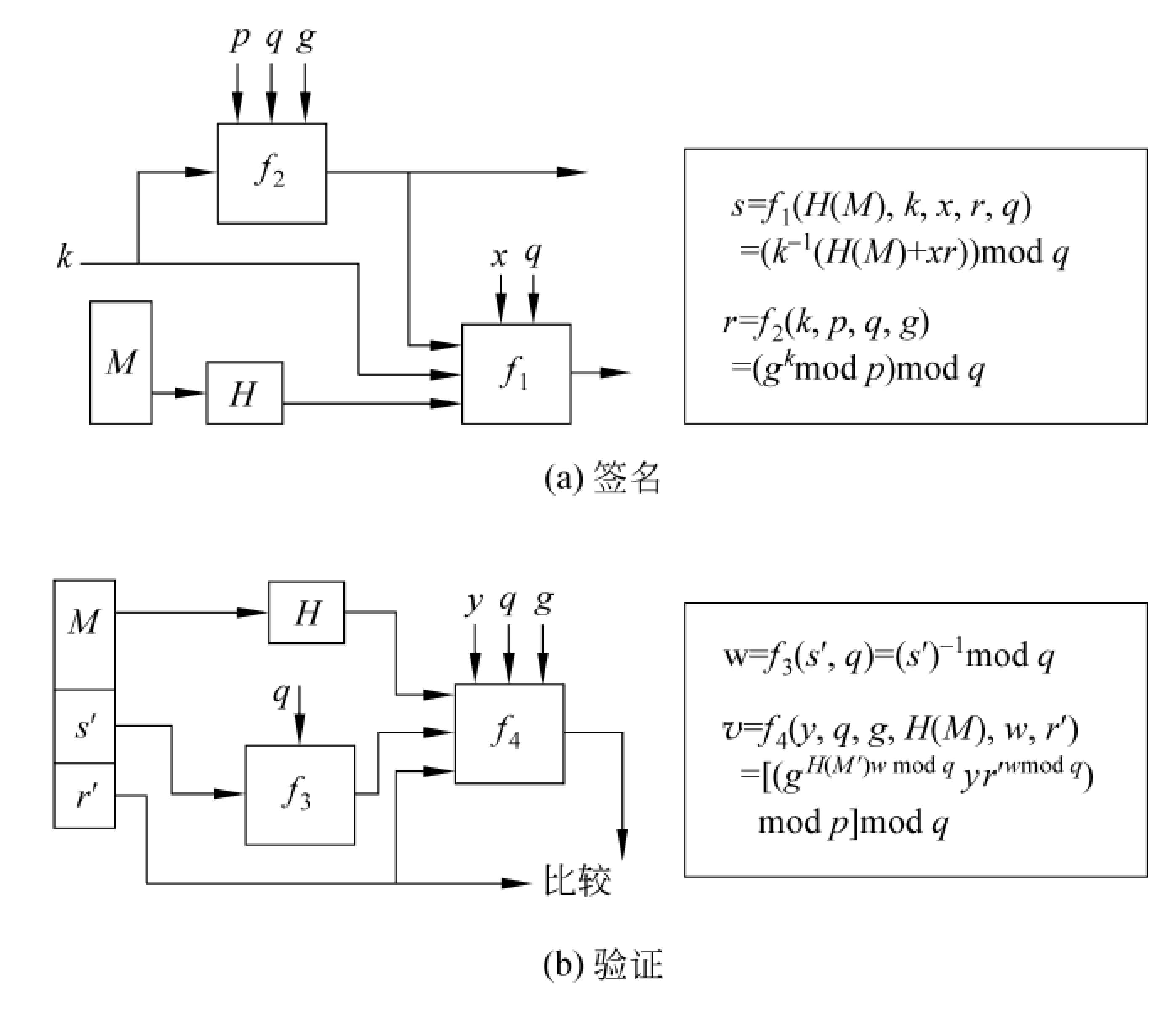


图 3-13 DSS 签名和验证

3.3.3 盲签名和群签名

1. 盲签名

盲签名的思想最早在 1982 年提出。盲签名因为具有盲性这一特点，可以有效保护所签署消息的具体内容，所以在电子商务和电子选举等领域有着广泛的应用。

盲签名允许使用者先将消息盲化，而后让签名者对盲化的消息进行签名，最后消息拥有者对签名除去盲因子，得到签名者关于原消息的签名。盲签名就是接收者在不让签名者获取所签署消息具体内容的情况下所采取的一种特殊的数字签名技术，它除了满足一般的数字签名条件外，还必须满足下面的两条性质：

(1) 签名者对其所签署的消息是不可见的，即签名者不知道他所签署消息的具体内容。

(2) 签名消息不可追踪，即当签名消息被公布后，签名者无法知道这是他哪次签署的。

曾经有人对盲签名给出了一个非常直观的说明：所谓盲签名，就是先将隐蔽的文件放进信封里，而除去盲因子的过程就是打开这个信封，当文件在一个信封中时，任何人不能读它。对文件签名就是通过在信封里放一张复写纸，签名者在信封上签名时，他的签名便透过复写纸签到文件上。

一般来说，一个好的盲签名应该具有以下的性质：

(1) 不可伪造性。除了签名者本人外，任何人都不能以他的名义生成有效的盲签名，这是一条最基本的性质。

(2) 不可抵赖性。签名者一旦签署了某个消息，就无法否认自己对消息的签名。

（3）盲性。签名者虽然对某个消息进行了签名，但他不可能得到消息的具体内容。

（4）不可跟踪性。一旦消息的签名公开后，签名者不能确定自己是在何时签署的这条消息。

满足上面几条性质的盲签名，被认为是安全的。这四条性质既是设计盲签名所应遵循的标准，又是判断盲签名性能优劣的根据。

另外，方案的可操作性和实现的效率也是设计盲签名时必须考虑的重要因素。一个盲签名的可操作性和实现速度取决于以下几个方面：

① 密钥的长度；

② 盲签名的长度；

③ 盲签名的算法和验证算法。

2. 群签名

群签名（group signature）是在1991年由Chaum和Van Heyst首次提出的一个签名概念。Camenish、Stadler、Tsudik等对这个概念进行了修改和完善。群签名在管理军事、政治及经济等多个方面有着广泛的应用。

群签名就是满足这样要求的签名：在一个群签名方案中，一个群体中的任意一个成员可以以匿名的方式代表整个群体对消息进行签名。与其他数字签名一样，群签名是可以公开验证的，而且可以只用单个群公钥来验证。也可以作为群标志来展示群的主要用途、种类等。

群签名技术主要经历了以下几个发展阶段。

（1）1991—1995年：在这段时间内，除了Chaum和Van Heyst给出的定义和四个实现群签名的方案外，主要是Chen和Pedersen的工作。Chen和Pedersen提出了几个新的群签名方案，同时首次提出了允许群体增加新成员的群签名方案。Camenish还对广义群签名进行了研究。

（2）1995—1997年：在经过几年对群签名的概念和意义的认识和理解之后，一些密码界人士开始对群签名技术进行研究。其间除了Chen和Pedersen的工作外，还有Park等的工作。在这一阶段，对群签名的研究不是十分活跃，主要是提出了一些新的群签名方案。

（3）1997年以后：自从1997年Camenish和Stadler首次提出适用于大的群体的群签名方案以来，群签名的研究进入了一个非常活跃的时期，取得了大量的研究成果。这些研究更注重群签名的安全性、效率和实用性，同时也涉及多个研究方向。这些研究有安全高效的群签名方案的研究，有群签名与通常的数字签名的相互转化的研究，还有群签名的推广方面的研究，如分级多群签名（group signatures for hierarchical multi-groups）、群盲签名（group blind signatures）、多群签名（multi-groups signatures）、子群签名（sub-group signatures）等，而且也取得了一些在电子商务方面的应用成果。因此Camenish和Stadler的研究成果已经成为群签名发展史上的一座里程碑。

群签名有下面几个研究方向：

（1）如何安全有效地废除群成员。即如何设计一个废除群成员的方法，使得一个群

成员被删除后,原来的私钥和成员证书不能再用于签名,而且不影响他原来所作的签名的安全性。现有的群签名方案都不能安全有效地废除群成员。

(2) 如何设计高效的打开签名的算法。即如何使群管理员不需要大的计算量就可以打开签名而确定出签名人的身份。

(3) 寻找一些安全高效的新的群签名算法。现有的相对安全高效的群签名方案基本上都依赖于RSA签名体制、Schnorr签名体制以及双重离散对数、离散对数的方根、有限循环群中元素的表示,某一秘密数值在一个指定的区间内的知识签名,效率都不是很高。因此,寻求新的安全高效的群签名算法是很有必要的。

(4) 如何在电子商务等领域更广泛地使用群签名。在现有的文献中,关于群签名在电子商务领域的应用还不多见。由于群签名对于签名人能提供良好的匿名性,同时又能使群管理员在必要的时候可以打开签名而撤销匿名性,所以可以广泛地应用于电子商务中的许多方面。只要能找到高效使用的群签名算法,群签名在电子商务中的应用必然会走向实用。

(5) 对于群签名相关的数字签名及其应用的研究。与群签名相关的数字签名及其应用的研究还不够。分级群签名、群盲签名、多群签名等都有实际应用背景,然而对它们的研究才处于起步阶段。

3.4 消息认证模式

3.4.1 消息的完整性与消息认证

消息完整性检验的一般机制如图3-14所示。无论是存储文件还是传输文件,都需要同时存储或发送该文件的数字指纹;验证时,对于实际得到的文件重新产生其数字指纹,再与原数字指纹进行对比,如果一致,则说明文件是完整的,即未被篡改、删除或插入,否则是不完整的。

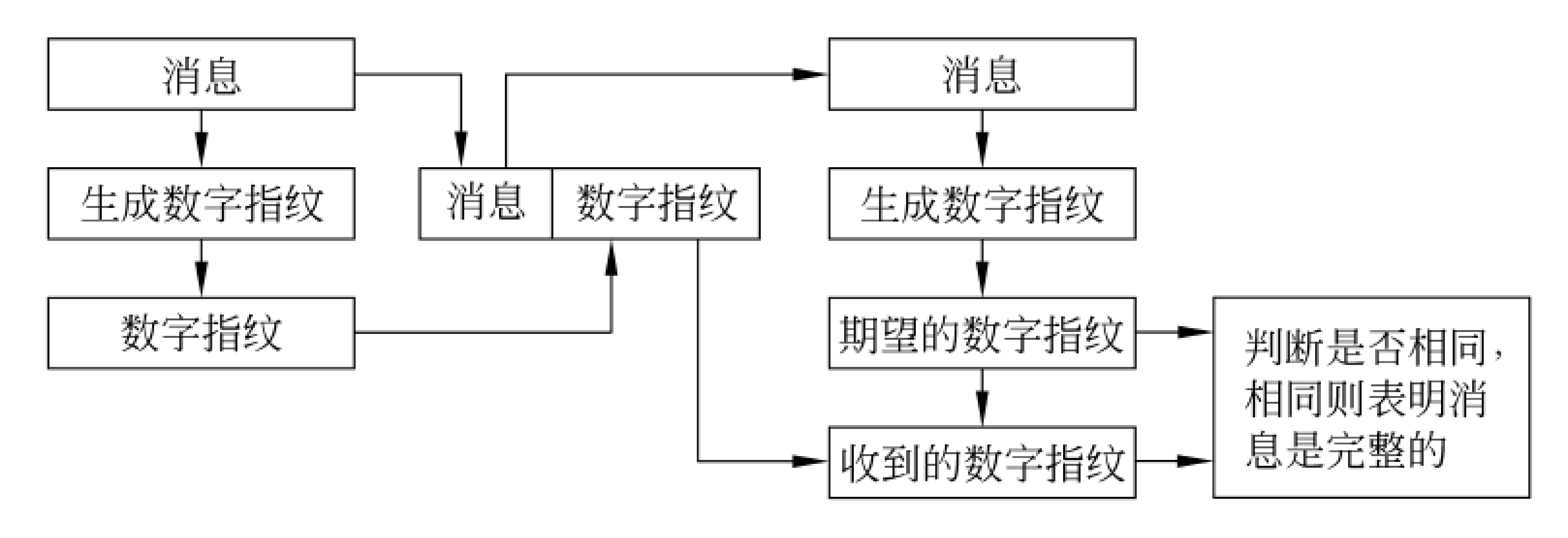


图3-14 消息完整性检验的一般机制

消息完整性检验只能检验消息是否完整,不能说明消息是否是伪造的。因为一个伪造的消息与其对应的数字指纹也是匹配的。

消息认证是指使一定的接收者能够检验接收到的消息是不是真实的方法。消息认证具有两层含义:一是检验消息的来源是否真实,即对消息的发送者的身份进行认证;二是检验消息的完整性,即验证消息在传送或存储过程中未被篡改、删除或插入等。

消息数字指纹的产生方法有很多。当需要进行消息认证时，仅有消息作为输入是不够的，需要加入密钥 K，这就是消息认证的原理。

消息认证码(Message Authentication Code，MAC)是与密钥相关的单向 Hash 函数。MAC 与单向 Hash 函数不同的是，它还包括一个密钥，不同的密钥会产生不同的 Hash 函数，这样就能在验证发送者的消息是否被篡改的同时，验证是由谁发送的。MAC 通常表示为：

$$\mathrm{MAC} = C_K(M)$$

其中，M 是长度可变的消息；K 是收、发双方共享的密钥；函数值 $C_K(M)$ 是定长的认证码，也称密码校验和。MAC 是带密钥的消息摘要函数，即一种带密钥的数字指纹，它与不带密钥的数字指纹是有本质区别的。

1. 消息认证

认证码被附加到消息后以 $M||\mathrm{MAC}$ 方式一并发送，接收方通过重新计算 MAC 以实现对 M 的认证，如图 3-15 所示。

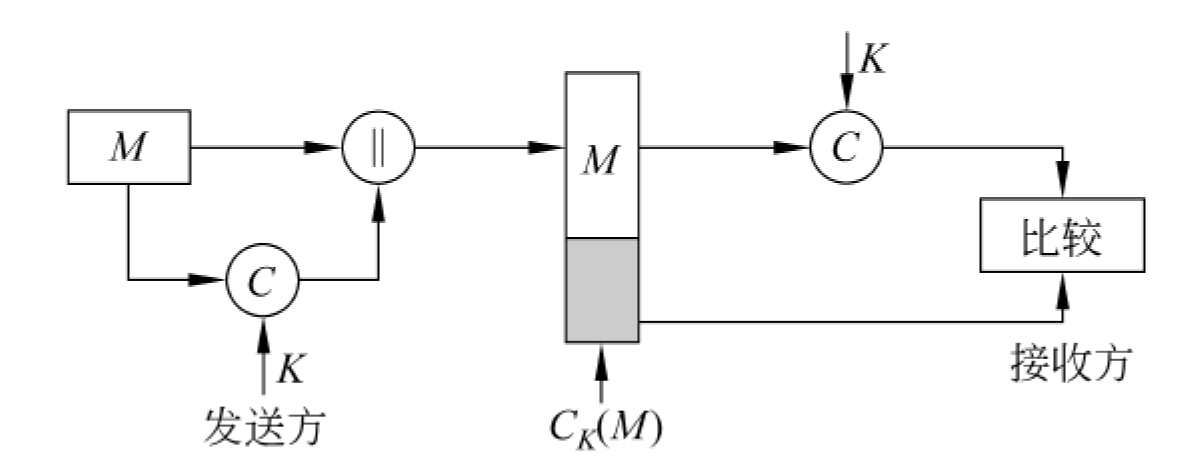


图 3-15　消息认证

假定发送方和接收方共享一个密钥 K，如果接收方收到的 MAC 与计算得出的 MAC 一致，那么可以得出如下结论：

(1) 接收方确信消息 M 未被篡改。此为完整性验证。

(2) 接收方确信消息来自所声称的发送者，因为没有其他人知道这个共享密钥，所以其他人也就不可能为消息 M 附加合适 MAC。此为消息源验证。

2. 消息认证与保密

在上述消息认证中，消息是以明文方式传送的，所以这一个过程只提供认证而不具备保密性。为提供保密性，可在 MAC 函数以后进行一次加密，而且加密密钥需被收、发双方共享，如图 3-16 所示。发送方发送 $E_{K_2}((M)||C_{K_1}(M))$。这种方式除具备认证的

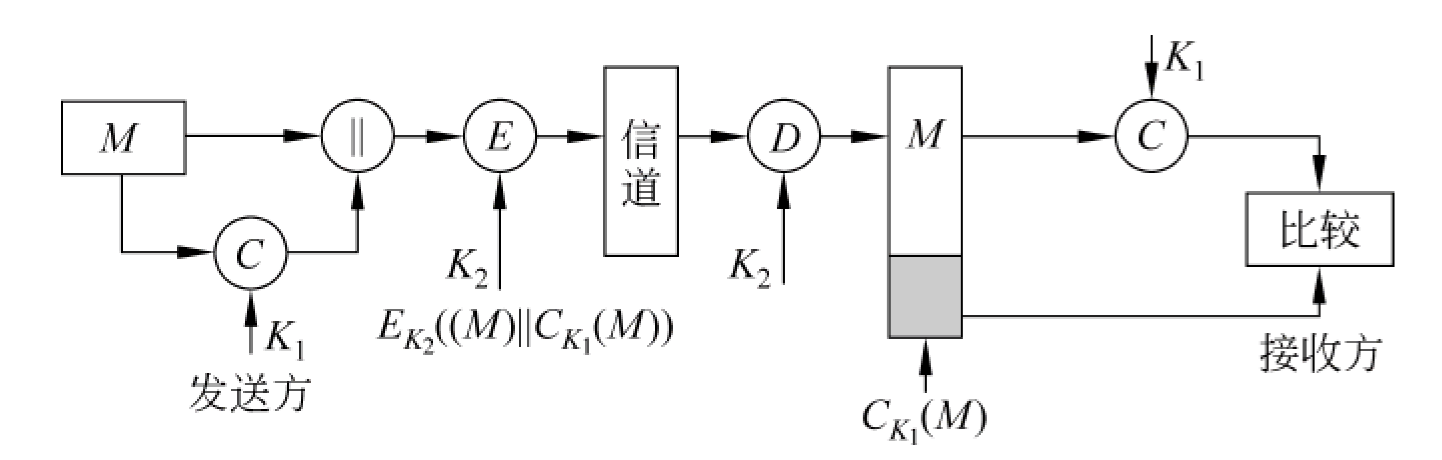


图 3-16　消息认证与保密

功能,还具有保密性。

3.4.2 消息认证模式

认证是认证者对被认证者的判定,认证活动又称为鉴别,是证明某人或对象身份的过程,是认证者对被认证者确定的过程。鉴定是对事务的提取与区分,是判定识别个体的过程,在现实世界,可以通过这个人的生物特征或他所拥有的某种物件,将他与其他个体区别开来,如 DNA 信息、虹膜、指纹、IC 卡等。例如,我们去乘坐飞机时,必须提供身份证或是驾驶证、军人证等进行自我证明,这就是鉴定。当公司的工作人员上班时,必须使用钥匙或非接触式 IC 门卡才能打开门锁进办公室的门,这是门锁对公司成员的认证,人们可以使用相同的钥匙来打开同一把锁(这就像数字世界中使用相同的用户名和口令),但是却不能判定具体是由哪一个公司成员打开的门锁。这就是认证和鉴定的不同。但有时对此区分得并不太严格,通常都是消息的接收者能够确认消息的来源或是判断授权用户是否能够访问网络。

通常,可以使用如下三种模式对网络认证加以保护:

(1) 用户认证:是基于使用者本身的认证。

(2) 会话认证:是对于用户访问服务权限的认证。

(3) 客户认证:一般基于源地址而不是基于用户的访问授权的认证。

3.4.3 消息认证方式

消息内容认证常用的方法是:消息发送者在消息中加入一个鉴别码(MAC、MDC 等)并经加密后发送给接受者(有时只需加密鉴别码即可)。接收者利用约定的算法对解密后的消息进行鉴别运算,将得到的鉴别码与收到的鉴别码进行比较,若二者相等,则接收,否则拒绝接收。

消息认证常用的三种方式为:

(1) 消息加密:用整个消息的密文作为认证标识。

(2) MAC:一个公开函数,加上一个密钥产生一个固定长度的值作为认证标识。

(3) Hash 函数:一个公开函数将任意长度的消息映射到一个固定长度的散列值,作为认证标识。

思 考 题

3.1 如何使用高级加密标准(AES)作为安全 Hash 函数(注意 Hash 函数不使用密钥。提示:可借鉴 Tiger 算法的外循环)?

3.2 假设要加密一个由三块分组明文 P_0、P_1 和 P_2 组成的消息,只用 Hash 函数和一个对称密钥 K,怎样对这个消息进行加密和解密?

3.3 对于除数为 10011 的 CRC 校验,找出数据 11010110,攻击者想要将数据改为 111*****,这里 * 表示对攻击者无关紧要的位。希望结果的校验值和原始数据的校验值

一致，找出所有能够选择的数据。

3.4 我们称变换 T 是增加的，如果它满足如下特性：对 M 进行某次 T 变换时，对于 M 的变换所需要的时间与前面所有对于 M 的变换所需的总时间成正比。假设有一个增加的 Hash 函数 H。

(1) 给出一种应用，使得增加的 Hash 函数 H 明显优于一般的（非增加的）Hash 函数。

(2) 假设消息 M 只能通过附加更多的位来修改，即修改后的消息 M' 是对于 X 满足 $M'=(M,X)$。给定一个安全的 Hash 函数 h，使用 h 定义一种增加的 Hash 函数 H。

3.5 找出下列两条消息（表示为十六进制）的所有不同的位，并验证它们的 MD5 Hash 值是相同的：

```
d1 31 dd 02 c5 e6 ee c4   69 3d 9a 06 98 af f9 5c
2f ca b5 87 12 46 7e ab   40 04 58 3e b8 fb 7f 89
55 ad 34 06 09 f4 b3 02   83 e4 88 83 25 71 41 5a
08 51 25 e8 f7 cd c9 9f   d9 1d bd f2 80 37 3c 5b
96 0b 1d d1 dc 41 7b 9c   e4 d8 97 f4 5a 65 55 d5
35 73 9a c7 f0 eb fd 0c   30 29 f1 66 d1 09 b1 8f
75 27 7f 79 30 d5 5c eb   22 e8 ad ba 79 cc 15 5c
ed 74 cb dd 5f c5 d3 6d   b1 9b 0a d8 35 cc a7 e3
```

和

```
d1 31 dd 02 c5 e6 ee c4   69 3d 9a 06 98 af f9 5c
2f ca b5 07 12 46 7e ab   40 04 58 3e b8 fb 7f 89
55 ad 34 06 09 f4 b3 02   83 e4 88 83 25 f1 41 5a
08 51 25 e8 f7 cd c9 9f   d9 1d bd 72 80 37 3c 5b
96 0b 1d d1 dc 41 7b 9c   e4 d8 97 f4 5a 65 55 d5
35 73 9a 47 f0 eb fd 0c   30 29 f1 66 d1 09 b1 8f
75 27 7f 79 30 d5 5c eb   22 e8 ad ba 79 4c 15 5c
ed 74 cb dd 5f c5 d3 6d   b1 9b 0a 58 35 cc a7 e3
```

3.6 因为 DSS 对每个签名产生一个 k，所以即使对同一消息签名，在不同的情况下签名也不同，单 RSA 签名则不能做到这一点。这种区别有什么实际意义？

3.7 可以利用 Hash 函数构造类似 DES 结构的分组密码。但 Hash 是单向的，而分组密码是可逆的（解密），那么如何用 Hash 码构造上述的分组密码呢？

3.8 如果用于产生 DSA 签名的 k 已被泄密，则会出现什么问题？

3.9 在 Diffle-Hellman 算法的基础上，设计可用于数字签名的方法是很有意义的。下面的方法比 DSA 更简单，它需要私钥但不需要秘密的随机数。

公开量：

$$q \text{ 为素数}$$

$$\alpha, \alpha<q \text{ 且 } \alpha \text{ 是 } q \text{ 的本原根}$$

私钥：$X, X<q$

公钥：$Y=\alpha^X \bmod q$

要对消息 M 签名，则先计算该消息的 Hash 码 $h=H(M)$。我们要求 $\gcd(h,q-1)=1$，若 $\gcd(h,q-1)$ 不为 1，则将 Hash 码附于消息后再计算 Hash 码，继续该过程直至生成的 Hash 码与 $(q-1)$ 互素，然后计算 $Zh\equiv X \bmod (q-1)$ 的 Z，并将 α^z 作为对该消息的签名。验证签名是验证 $Y=(\alpha^Z)^h=\alpha^X \bmod q$。

(1) 证明该体制能正确运行，即证明如果签名是有效的，那么验证过程中将有上述等式成立。

(2) 给出一种对给定的消息伪造用户签名的简单方法，以证明这种体制是不可接受的。

参考文献

[1] Jueneman R，Matyas S，Meyer C. Message Authentication. IEEE Communications Magazine，1988.

[2] Juneman R. Electronic Document Authentication. IEEE Network，1987，1(2)：17-23.

[3] Menezes A，Oorshcot P，Vanstone S. Handbook of Applied Cryptography. Boca Raton，FL：CRC Press，1997.

[4] Stinson D. Cryptography：Theory and Practice. Boca Raton，FL：CRC Press，2002.

[5] Black J，Rogaway P A. Block-Cipher Mode of Operation for Parallelizable Message Authentication. Advances in Cryptology-EUROCRYPT 2002，Heidelberg：Springer-Verlag，2002：384-401.

[6] Virgil D G，Pompiliu D. Fast Encryption and Authentication：XCBC Encryption and XECB Authentication Modes. In FSE 2001 Yokohama Heidelberg：Springer-Verlag，2002：92-141.

[7] Hugo Krawczyk. How Secure is SSL. CRYPTO 2001，Heidelberg：Springer-Verlag，2001：310-331.

[8] Rogaway P，Bellare M，Black J. OCB：A Block-Cipher Mode of Operation for Efficient Authenticated Encryption. ACM Conference on Computer and Communications Security (CCS'01)，Philadelphia，PA，USA，2001，ACMPress，2001：196-205.

[9] Deepakumara J，Heys H M，Venkatesan R. Performance comparison of message authentication code (MAC) algorithms for Internet protocol security (IPSEC). Proc. Newfoundland Electrical and Computer Engineering Conf.，St. John's，Newfoundland，Nov. 2003.

[10] Liskov M，Rivest R，Wagner D. Tweakable Block Ciphers，Advances in Cryptology. CRYPTO 2002 Heidelberg：Springer-Verlag，2002：31-46.

[11] Akl S G. Digital Signatures：A Tutorial Survey. Computer，1983，16(2)：15-24.

[12] Piper F，Blake-Wilson S，Mitchell J. Digital Signatures：Security and Control. Information System Audit and Control Foundation，2000.

[13] 王大印，林东岱，吴文玲. 消息认证码研究. 通讯和计算机，2005，2(10)：76-81.

[14] 王海艳，王汝佳. 群签名方案之比较研究. 计算机应用研究，2005，22(10)：93-95.

[15] 祝建华. 安全群签名体制研究及安全性分析. 华中科技大学博士学位论文，2009.

第 4 章 chapter 4

信息隐藏与数字水印

本章学习目标

信息隐藏与数字水印是当前数字内容安全的热门技术之一。本章将介绍信息隐藏与数字水印的基本原理，主要包括：信息隐藏与数字水印技术的基本概念、空域和变换域的信息隐藏技术、数字水印技术以及信息隐藏与数字水印的发展与应用等。

通过本章的学习，应掌握以下内容：

(1) 空域和变换域信息隐藏技术。

(2) 数字水印框架与评价指标。

(3) 版权保护与内容认证数字水印技术。

(4) 可逆水印技术。

(5) 信息隐藏与数字水印技术的发展历程与应用情况。

4.1 基本概念

最早的隐写术可以追溯到远古年代。

(1) 用头发掩盖信息：大约在公元前 440 年，为了鼓励奴隶们起来反抗，Histiaus 给他最信任的奴隶剃头，并将消息刺在头上，等到头发长出来后，消息被遮盖，这样消息就可以在各个部落中传递。

(2) 使用书记板隐藏信息：在波斯朝廷的一个希腊人 Demeratus，他要警告斯巴达将有一场由波斯国王薛西斯一世发动的入侵，他首先去掉书记板上的蜡，然后将消息写在木板上，再用蜡覆盖，这样处理后的书记板看起来是完全空白的。事实上，它几乎既欺骗了检查的士兵也欺骗了接受信息的人。

(3) 使用音乐谱隐藏信息：Schott(1608—1666)的 400 页的著作 *Schola Steganographica* 中，他阐述了如何在音乐乐谱中隐藏消息：每个音符对应于一个字符。Bach 提出了另一种基于音符的出现次数的方法。Schott 还扩展了 Trithemius(1462—1516)在 *Steganographice* 一书(这是有关这个领域的最早的一本著作)中提出的 Ave Maria 码。扩展码使用 40 个表，每个表有 24 个入口(当时，每个入口对应于字母表中的一个字母)，这些入口包括四种语言：拉丁文、德文、意大利文和法文。纯文本中的每个字母，被相应

入口内的词或短语所替代，最终隐秘文本看上去像是祈祷词或者咒语。最近的研究表明，通过把这些表对 25 取模并应用到一个逆转的字母表中，就可以破译它们。剑桥 Trinity 学院的教师 Wilkins(1614—1672)论述了“两个音乐家能够通过使用他们的乐器交谈，就像用嘴说话一样”是因为什么。他还解释了如何在几何图形中通过使用点、线和三角形来隐藏消息。“点、线段的终端和图的角度，都可以表示不同的字母”。

(4) 使用离合诗隐藏信息。在 Kahn 的 *The Codebreakers* 一书中，他例举了一个修道士是如何写下一本书并把他的心上人的名字设为连续章节的第一个字母。他还例举了一个战俘如何在寄回家的信中隐藏消息，这个战俘在 I、J、T、F 上使用点和虚线，用来拼写一条 Morse 编码的隐藏文本。这些“语义方法”隐藏了消息，但是却有一个内在的问题，掩饰文本难于创建，并且常常听起来很奇怪，这足以引起保密检查员的注意。在两次世界大战中，保密检查员截获了大量这样的消息。一个著名的例子是：第一次世界大战中，一份海底电报说“父亲去世了”，保密检查员将它修改为“父亲病了”并发送出去。对这份修改过的电报回复泄露了秘密：“父亲去世了，还是病了？”

(5) 使用微小图隐藏信息：在 1857 年，Brewster 建议隐藏保密消息到“那些不超过句号或者一小滴墨水的空间”中。在 1870—1871 年爆发的法国和普鲁士的战争中，当巴黎被围困时，鸽子带出了隐藏在微缩胶卷上的消息。在 1905 年的俄国和日本的战争中，显微图像被隐藏于耳朵、鼻孔中，甚至指甲之下。在第一次世界大战中，间谍们收发的消息通过几次照相缩小成为细小的点，然后把这些点粘贴在那些无关紧要的掩饰材料如杂志中印刷的逗号之上。

多媒体数据的数字化为多媒体信息的存取提供了极大的便利，同时也极大地提高了信息表达的效率和准确性。随着因特网的日益普及，多媒体信息的交流已达到了前所未有的深度和广度，其发布形式也愈加丰富了。人们如今也可以通过因特网发布自己的作品、重要信息和进行网络贸易等 ，但是随之而出现的问题也十分严重：如作品侵权更加容易，篡改也更加方便。因此如何既充分利用因特网的便利 ，又能有效地保护知识产权，已受到人们的高度重视。这标志着一门新兴的交叉学科——信息隐藏学的正式诞生。如今信息隐藏学作为隐蔽通信和知识产权保护等的主要手段，正得到广泛的研究与应用。本章从信息隐藏和数字水印的基本原理、常用模型入手，然后对这些技术的应用进行详细的介绍。

信息隐藏(information hiding)，也叫数据隐藏(data hiding)。简单地说，信息隐藏就是将秘密信息隐藏于另一非保密的载体之中。这里的载体可以是图像、音频、视频、文本，也可以是信道，甚至是某套编码体制或整个系统。

信息之所以能够隐藏在多媒体数据中主要是基于两个事实。其一，多媒体信息本身存在很大的冗余性。从信息论的角度看，未压缩的多媒体信息的编码效率是很低的，所以将这些机密信息嵌入到多媒体信息中进行秘密传送是完全可行的，并不会影响到多媒体信息本身的传送和使用。其二，人类的听觉和视觉系统都有一定的掩蔽效应。人们可以充分利用这种掩蔽性将信息隐藏而不被察觉。

信息隐藏不同于传统的密码学技术。密码技术主要是研究如何将机密信息进行特殊的编码，以形成不可识别的密码形式(密文)进行传递；而信息隐藏则主要研究如何将

某一机密信息秘密隐藏于另一公开的信息中，然后通过公开信息的传输来传递机密信息。对加密通信而言，可能的监测者或非法拦截者可通过截取密文，并对其进行破译，或将密文进行破坏后再发送，从而影响机密信息的安全；但对信息隐藏而言，可能的监测者或非法拦截者则难以从公开信息中判断机密信息是否存在，难以截获机密信息，从而能保证机密信息的安全。多媒体技术的广泛应用，为信息隐藏技术的发展提供了更加广阔的领域。

随着数字技术和 Internet 的发展，各种形式的多媒体数字作品（图像、视频、音频等）纷纷以网络形式发表，其版权保护成为一个迫切需要解决的问题。由于数字水印（digital watermarking）是实现版权保护的有效办法，因此如今已成为多媒体信息安全研究领域的一个热点，也是信息隐藏技术研究领域的重要分支。该技术即是通过在原始数据中嵌入秘密信息——水印（watermark）来证实该数据的所有权。这种被嵌入的水印可以是一段文字、标识、序列号等，而且这种水印通常是不可见或不可察的，它与原始数据（如图像、音频、视频数据）紧密结合并隐藏其中，并可以经历一些不破坏原数据使用价值或商用价值的操作而能保存下来。数字水印技术除了应具备信息隐藏技术的一般特点外，还有着其固有的特点和研究方法 。在数字水印系统中，隐藏信息的丢失，即意味着版权信息的丢失，从而也就失去了版权保护的功能，也就是说，这一系统就是失败的。由此可见，数字水印技术必须具有较强的鲁棒性、安全性和透明性。

下面给出一些基本的定义。对于通信的双方 A 和 B，A 希望将秘密传递给 B，A 需要从一些随机消息源中选取一个消息 h，这个消息在公开传递时不会引起怀疑，我们称 h 为载体对象。然后把需要传递的秘密信息 m 隐藏到载体对象 h 中，这样，载体对象 h 就变成了伪装对象 h'。伪装对象和载体对象在感官效果（包括视觉、听觉等）上是不可区分的。这样就实现了信息的隐秘传递，它掩盖了信息传输的事实，实现了信息的安全传递。

秘密信息在嵌入过程中，可能需要密钥，也可能不需要密钥。这里，为了区别密码中的密钥，信息隐藏的密钥通常称为伪装密钥。

图 4-1 即为信息隐藏的原理框图。A 首先从载体信息源中选择一个载体信号，采用信息嵌入算法将密码信息 m 嵌入载体信号中，嵌入算法可能会用到密钥。嵌入了信息的载体通过公开信道传递给 B。用户 B 接收到信息后，由于他知道 A 使用的嵌入算法和嵌入密钥，他可以利用相应的提取算法将隐藏在载体中的秘密信息提取出来。提出过程中可能需要（或不需要）原始载体对象 h，这取决于具体所使用的信息嵌入算法。

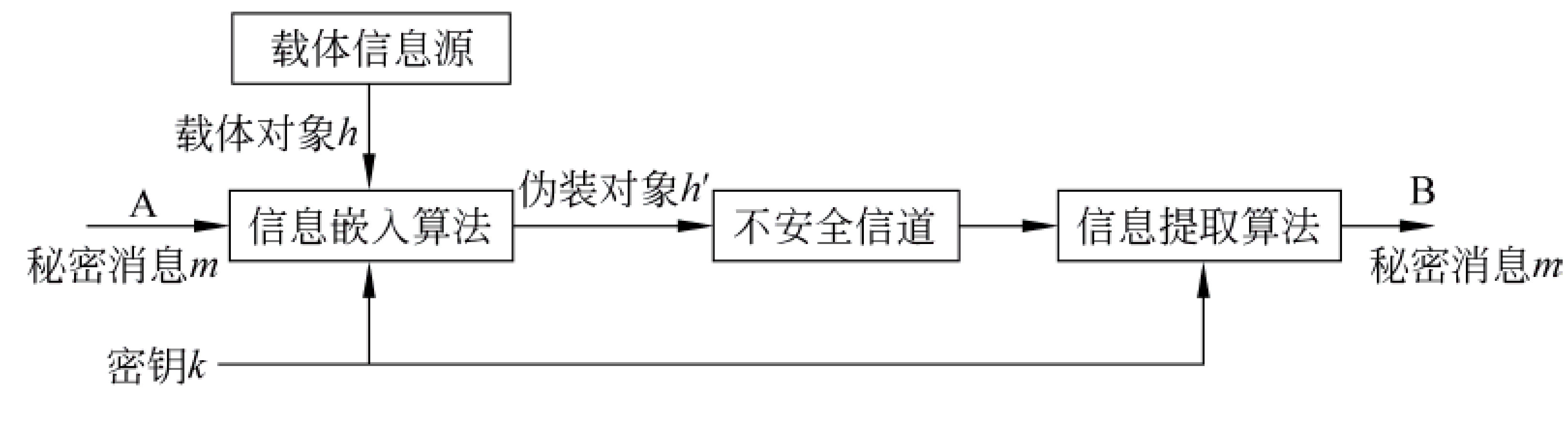


图 4-1 信息隐藏原理框图

根据信息隐藏技术的应用目的和载体对象不同，信息隐藏可分为许多分支。下面对几个主要分支隐写术、数字水印、阈下信道作简单介绍。

(1) 隐写术。隐写术(steganogeaphy)这个术语来自希腊词汇 steganos 和 graphia,前者的含义是“秘密的”,后者的含义是“书写”。隐写术是一种隐蔽通信技术,其主要目的是将重要的信息隐藏起来,以便不引起人注意地进行传输和存储。隐写术在其发展过程中逐渐形成了两大分支,即语义隐写和技术隐写。

语义隐写术利用了语言文字自身及其修辞方面的知识和技巧,通过对原文按照一定规则进行重新排列或剪裁,从而隐藏和提取密文。语义隐写术包括符号码、隐语以及虚字密码等。所谓符号码是指一次非书面形式的秘密通信。例如,第二次世界大战中,有人曾经利用一幅关于圣安东尼奥河的画传递了一封密信。画中的圣安东尼奥河畔长了许多小草,而小草的叶子的长短是根据一种编码画出来的。长叶代表莫尔斯电码的划线,短叶代表莫尔斯电码的圆点。得到这幅画的人利用电码本很容易就得到了信的内容。这是符号码成功应用的一个典范。另一个例子是,在第二次世界大战中,检查者截获了一批手表,由于担心手表的指针位置会拼出一条秘密消息,他们在检查过程中对指针的位置进行了调整。这种利用手表指针位置传递秘密消息也属于符号码类型的语义隐写术。需要注意的是,符号码的结果不能影响载体的特征,例如上述画中的草叶的形状和分布必须符合常规,否则就是隐写失败。隐语所利用的是错觉或代码字。在第一次世界大战中,德国间谍曾使用雪茄的假定订单来代表不同类型的英国军舰——巡洋舰和驱逐舰,例如朴茨茅斯需要 5000 根雪茄就代表着朴茨茅斯有 5 艘巡洋舰。另外,在第二次世界大战期间,一个名叫 Valer Dickinson 的妇女使用玩偶作为代码字表示美国在纽约的船只数目来向日本发送信息,她是用小玩偶代表驱逐舰,而用大玩偶代表航空母舰或战舰。在虚字密码中通常是使用每个单词的相同位置的字母来拼出一条消息。但是这样的载体消息非常难以构造。我国古代经常出现的“藏头诗”就是一种典型的虚字密码形式。

技术隐写术是隐写术中的主要分支。毫无疑问,技术隐写术的发展是伴随着科技,尤其是信息科技的发展而发展的。从古代的利用动物的身体记载、木片上打蜡,到近代使用的隐形墨水、缩微胶片,再到当代使用的扩频通信、网络多媒体数据隐写等,可以说每一种新隐写术的出现都离不开科学技术的进步。当代出现的与数字载体有关的隐写术都可以借鉴到数字水印的技术领域,故在此不做展开。

(2) 数字水印。数字水印技术是信息隐藏技术的另一重要分支,其基本思想是在数字作品(图像、音频、视频等)中嵌入秘密信息,以便保护数字产品的版权、证明产品的真实可靠性、跟踪盗版行为或提供产品的附加信息。其中的秘密信息可以是版权标志、用户序列号或者产品相关信息。数字水印是本章讨论的重点,留待后面详细介绍,这里简单说明一下数字水印与隐写术的区别。

隐写术和数字水印的基本思想都是将秘密信息隐藏在载体对象中。但是两者之间还是有本质的不同的。在隐写术应用中,所要发送的秘密信息是主体,是重点保护对象,而用什么载体对对象进行传输无关紧要。对于数字水印来说,载体通常是数字产品,是版权保护对象,而所嵌入的信息则是与该产品相关的版权标志或相关信息。

(3) 隐蔽信道。隐蔽信道是指允许进程以危害系统安全策略的方式传输信息的通信信道。我国的《计算机信息系统安全保护等级划分准则》(GB 17859—1999)、美国的《可

信计算机系统评估准则》(TCSEC)以及国际标准化组织ISO在1999年发布的《信息技术安全评估通用准则》(ISO/IEC 15408,简称CC标准)都对隐蔽信道分析提出了明确的规定。要求高等级信息系统(GB 17859—1999第四级,TCSEC中B2级以上)必须进行隐蔽信道分析,在识别隐蔽信道的基础上,对隐蔽信道进行度量和处置。

隐蔽信道的概念最初是由Lampson在1973年提出的,其给出的隐蔽信道定义为:不是被设计或本意不是用来传输信息的通信信道。Lampson关注程序的限制问题,即如何在程序的执行过程中进行限制,使其不能向其他未授权的程序传输信息。他列举了恶意或行为不当的程序绕过限制措施,泄露数据的6种方法和相应的处理措施,并把这些方法归纳为3种类型:存储信道、合法信道和隐蔽信道。后续的研究将隐蔽信道重新划分为两种类型:存储隐蔽信道和时间隐蔽信道,统称隐蔽信道。其中,时间隐蔽信道对应于Lampson所指的"隐蔽信道";合法信道则是一种阈下信道(subliminal channel),是公开信道中所建立的一种实现隐蔽通信的方式。信道中公开的、有意义的信息仅仅充当了秘密信息的载体,秘密信息通过它进行传输。这种隐蔽传输信息的方式后来逐渐淡出了隐蔽信道研究的中心,形成了相对独立的研究领域。

隐蔽信道分析工作包括信道识别、度量和处置。信道识别是对系统的静态分析,强调对设计和代码进行分析发现所有潜在的隐蔽信道。信道度量是对信道传输能力和威胁程度的评价。信道处置措施包括信道消除、限制和审计。隐蔽信道消除措施包括修改系统、排除产生隐蔽信道的源头、破坏信道的存在条件。限制措施要求将信道危害降低到系统能够容忍的范围内。但是,并非所有的潜在隐蔽信道都能被入侵者实际利用,如果对所有的潜在隐蔽信道进行度量和处置会产生不必要的性能消耗,降低系统效率。隐蔽信道检测则强调对潜在隐蔽信道的相关操作进行监测和记录,通过分析记录,检测出入侵者对信道的实际使用操作,为信道度量和处置提供依据。

(4) 阈下信道。阈下信道是指在基于公钥密码技术的数字签名、认证等应用密码体制的输出密码数据中建立起来的一种隐蔽信道,除指定的接收者外,任何其他人均不知道密码数据中是否有阈下消息存在。

Gustavus Simmons发明了传统数字签名算法中阈下信道的概念。由于阈下信道隐藏在看似正常的数字签名的文本中,所以这是一种迷惑人的信息传递。事实上,阈下信道签名算法与通常的签名算法区别不开,至少对Walter是这样,Walter不仅读不出阈下信道消息,而且他也不知道阈下信道已经出现。

1983年,Simmons把隐蔽通信问题表述为"囚犯问题",如图4-2所示。在该模型中,

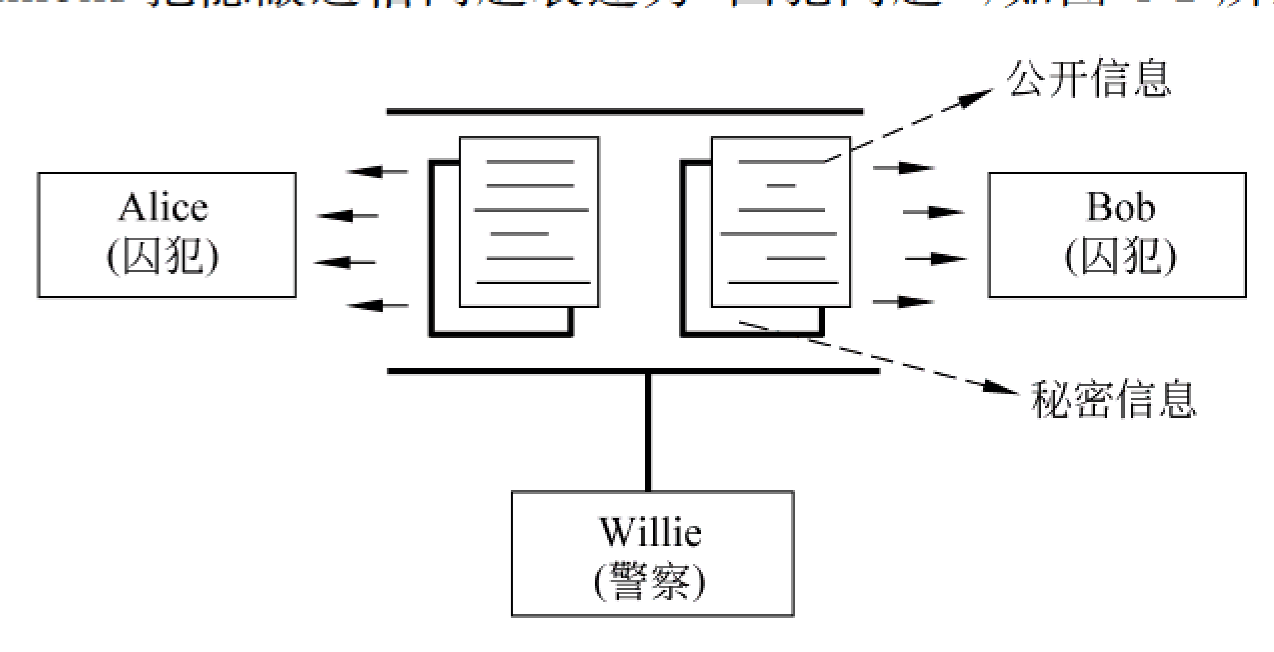


图4-2 "囚犯问题"模型

囚犯 Alice 和 Bob 被关押在监狱的不同牢房里，他们准备越狱，故需通过一种隐蔽的方式交换信息，但他们之间的通信必须通过狱警 Willie 的检查。因此，他们必须找到一种办法，可以将秘密的信息隐藏在普通的信息里。

囚犯问题根据 Willie 的反应方式分为被动狱警问题、主动狱警问题及恶意狱警问题 3 种。

① 被动狱警问题：狱警 Willie 只检查他们之间传递的信息有没有可疑的地方，一旦发现有可疑信息甚至是非法信息通过，就会立即做出相应的反应。

② 主动狱警问题：狱警 Willie 在不破坏公开信息的前提下，故意去修改一些可能隐藏有机密信息的地方，以达到破坏可能的机密信息的目的。例如，对于文本数据，他可能会把其中一些词句用相近的同义词来代替，而不改变通信内容。

③ 恶意狱警问题：狱警 Willie 可能彻底改变通信囚犯的信息，或者伪装成一个囚犯，隐藏伪造的机密信息，发给另外的囚犯。在这种条件下，囚犯可能就会上当，他的真实想法就会暴露无遗。对这种情况，囚犯是无能为力的。不过现实生活中，这种恶意破坏通信内容的行为一般是不允许的，有诱骗嫌疑。目前的研究工作重点是针对主动狱警问题。

一个简单的阈下信道可以是句子中单词的数目。句子中奇数个单词对应 1，而偶数个单词对应 0。因此，当读这种仿佛无关紧要的句子时，已经将信息 1010 传递给了自己的接收人员。不过这个例子的问题在于它没有密钥，安全性完全依赖于算法的保密性。

4.2 信息隐藏技术

4.2.1 信息隐藏技术的发展历程

1992 年，国际上正式提出信息隐藏的概念；1996 年，在英国剑桥大学牛顿研究所召开了第一届信息隐藏学术会议，标志着信息隐藏学的正式诞生。此后，国际信息隐藏学术会议在欧美各国相继召开，至今已举办十三届之多。

作为隐秘通信和知识产权保护等的主要手段，信息隐藏从正式提出到现在二十年的时间里引起了各国政府、大学和研究机构的重视，取得了巨大的发展。美国的麻省理工学院、普渡大学，英国的剑桥大学、NEC 研究所、IBM 研究所都进行了大量的研究。在国内，许多高等院校和研究机构也对信息隐藏技术进行了深入的研究。从 1999 年开始，我国已召开了九届全国性的信息隐藏暨多媒体内容安全学术大会。国家 863 计划智能计算机专家组会同中国科学院自动化研究所模式识别国家重点实验室和北京邮电大学信息安全中心还召开了专门的“数字水印学术研讨会”。

随着理论研究的进行，相关的应用技术和软件也不断推出。如美国 Digimarc 公司在 1995 年开发了水印制作技术，是当时世界上唯一一家拥有这一技术的公司，并在 Photoshop 4.0 和 CoreDraw 7.0 中进行了应用。日本电器公司、日立制作所、先锋、索尼和 IBM 公司在 1999 年宣布联合开发统一标准的基于数字水印技术的 DVD 影碟防盗版技术。DVD 影碟在理论上可以无限制地复制高质量的画面和声音，因此迫切需要有效

的防盗版技术。该技术的应用使消费者可以复制高质量的动态图像，但以赢利为目的的大批量非法复制则无法进行。2000 年，德国在数字水印保护和防止伪造电子照片的技术方面取得了突破。以制作个人身份证为例，一般要经过扫描照片和签名、输入制证机、打印和塑封等过程。

上述新技术是在打印证件前，在照片上附加一个暗藏的数字水印。具体做法是在照片上对某些不为人注意的部分进行改动，处理后的照片用肉眼看与原来几乎一样，只有用专用的扫描器才能发现水印，从而可以迅速、无误地确定证件的真伪。该系统既可在照片上加上牢固的水印，也可以经改动使水印消失，使任何伪造企图都无法得逞。由欧盟委员会资助的几个国际研究项目也正致力于实用的水印技术研究，欧盟期望能使其成员国在数字作品电子交易方面达成协议，其中的数字水印系统可以提供对复制品的探测追踪。在数字作品转让之前，作品创作者可以嵌入创作标志水印；作品转让后，媒体发行者对存储在服务器中的作品加入发行者标志；在出售作品拷贝时，还要加入销售标志。

经过多年的努力，信息隐藏技术的研究已经取得了许多成果。从技术上来看，隐藏有机密信息的载体不但能经受人的感觉检测和仪器设备的检测，而且还能抵抗各种人为的蓄意攻击。但总的来说，信息隐藏技术尚未发展到可大规模使用的阶段，仍有不少理论和技术性的问题需要解决。到目前为止，信息隐藏技术还没有形成自身的理论体系。例如，如何计算一个数字媒体或文件所能隐藏的最大安全信息量等。尽管信息隐藏技术在理论研究、技术开发和实用性方面尚不成熟，但它的特殊作用，特别是在数字版权保护方面的独特作用，可以说是任何其他技术无法取代的，我们有理由相信信息隐藏技术必将在未来的信息安全体系中独树一帜。

信息隐藏的目的在于把机密信息隐藏于可以公开的信息载体之中，信息载体可以是任何一种多媒体数据，如音频、视频、图像，甚至文本数据等，被隐藏的机密信息也可以是任何形式。一个很自然的要求是，信息隐藏后能够防止第三方从信息载体中获取或检测出机密信息。

4.2.2 信息隐藏技术的分类与要求

根据应用场合的不同要求，信息隐藏技术可以分为隐写术和数字水印两个主要分支。隐写术研究的重点是如何实现信息伪装的隐蔽性；而数字水印则需要考虑水印信息是否稳健等特性，如对各种可能攻击的敏感性等。根据隐藏协议，信息隐藏还可分为无密钥信息隐藏、私钥信息隐藏、公钥信息隐藏。

数字水印近年来受到了信息隐藏研究人员的广泛关注。水印可以是标注版权的信息或 ID、图形或图章、音频信息、随机序列等。数字水印根据宿主信息的不同，可分为文本水印、图像水印、视频水印、矢量图水印等。图像、语音、视频信号通常具有较大的感官冗余，故能提供较大的信息隐藏空间。

根据水印嵌入所处的位置，水印可分为空域数字水印和变换域数字水印。根据数字水印的性质，水印可以分为鲁棒水印(robust watermarks)和脆弱水印(fragile watermarks)。两类水印的用途完全不同。鲁棒水印主要用于数字内容信息的版权保护和所有权认定，故应能经受各种潜在的攻击；脆弱水印可以进一步分为完全脆弱水印和半脆弱水印

(semi-fragile watermarks)。完全脆弱水印对任何针对含水印载体的处理都非常敏感，而半脆弱水印则只对恶意的处理敏感，而对合法的处理鲁棒。在实际应用中，半脆弱水印通常具有更广泛的应用前景。

根据水印检测是否需要原始载体信息和原始水印信息，数字水印可以分为盲检测水印(blind detection)和非盲检测水印。从检测方法的角度，水印可以分为私有水印(private watermark)和公开水印(public watermark)。此外，根据含水印载体是否可无损恢复，水印还可分为可逆水印(reversible watermark)和不可逆水印(irreversible watermark)。

不同的应用场合需要采用不同的信息隐藏技术，它们的要求也不同。

(1) 隐写术：对隐写术最重要的要求包括不可感知性和不可检测性、秘密性、较大的水印容量以及算法实现简单。

(2) 鲁棒水印：对鲁棒水印最重要的要求包括不可感知性、鲁棒性(即含水印的载体经过一些信号处理以后，水印仍然具有较好的可检测性)、能解决所有权死锁问题、秘密性以及算法实现简单等。

(3) 完全脆弱水印：对完全脆弱水印最重要的要求包括不可感知性、对任何处理的敏感性、秘密性以及算法实现简单等。

(4) 半脆弱水印：对完全脆弱水印最重要的要求包括不可感知性、对恶意攻击的敏感性、对合法处理的鲁棒性、秘密性以及算法实现简单等。

在信息隐藏中，三个最主要的因素分别是鲁棒性、不可感知性和水印容量。在上述三个因素的关系上，J. Fridich 给出了如图 4-3 所示的三角关系。它的含义是：对于一个信息隐藏系统，在这三个要素上总是会在某一个上有所偏重，不可能同时达到最优。例如，如果我们希望一个信息隐藏系统的鲁棒性很好，那就会不可避免地在水印容量和不可感知性方面做出一定的牺牲。

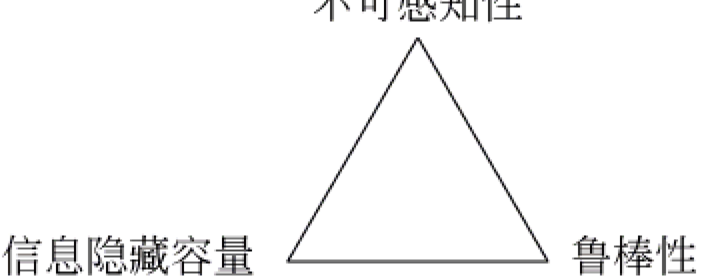


图 4-3　信息隐藏三个要素的关系

4.2.3　信息隐藏技术的基本原理与模型

从信号处理的角度来理解，信息隐藏可视为在强背景信号(载体)中叠加一个弱信号(隐藏信息)。由于人的听觉系统和视觉系统的分辨能力受到一定的限制，叠加的弱信号只要低于某一个阈值，人就无法感觉到隐藏信息的存在。

设 H 和 H' 分别表示原始载体信号和隐藏信息后的含隐秘信息载体信号，W 为待隐藏信息，信息隐藏的过程可表示为：

$$H' = H + f(F, W) \tag{4-1}$$

I. J. Cox 提出了三种常用的信息嵌入公式，分别为：

$$h'_i = h_i + \alpha w_i \tag{4-2}$$

$$h'_i = h_i(1 + \alpha w_i) \tag{4-3}$$

$$h'_i = h_i + \alpha \mid h_i \mid w_i \tag{4-4}$$

其中，h_i 和 h'_i 分别表示原始载体信号和隐藏信息后的含隐秘信息载体信号分量(或从中

提取的特征)值，w_i 为待嵌入隐藏信号分量，α 为嵌入强度。α 越大，嵌入的信号幅度越大，鲁棒性越好，但感知性会降低；反之，则感知性好而鲁棒性降低。因此，α 的选择应在满足不可感知性的前提下，尽可能选择较大的值。

对于式(4-2)和式(4-3)所示的嵌入方法，可以实现盲检测。由于式(4-4)中 h_i 的符号改变的随机性，无法实现盲检测。

假设用 H^* 表示待测的掩密信号，从中提取的水印序列用 W^* 表示，$W^*=\{w_i^*\}$，在 H^* 相对于 H' 没有误差的情况下，隐藏信息可由式(4-2)和式(4-3)提取：

$$w_i^* = (h_i^* - h_i)/\alpha \quad \text{或} \quad w_i^* = (h_i^* - h_i)/\alpha \cdot h_i$$

然而，由于 H^* 相对于 H' 会有一些失真，因此提取出来的 w_i^* 也会和原始的隐藏信息 w_i 不同。为此，水印的检测通常需要三个步骤：

(1) 计算检测的水印与原始水印信息的相关性。

(2) 门限化所得到的计算结果。

(3) 判断水印是否存在。

为了确定 H^* 中是否含有水印，可以通过式(4-5)计算 W^* 和 W 的相似度：

$$\rho(W^*, W) = \sum_{i=0}^{K-1} w_i^* w_i \Big/ \sqrt{\sum_{i=0}^{K-1} (w_i^*)^2} \tag{4-5}$$

水印存在与否的判定标准为：若 $\rho(W^*, W) > T$，可以判定被测掩密信号中有水印 W 存在；否则没有。T 为一阈值，其选择需要综合考虑误检率和漏检率。T 值选择过小，会导致误检率增加而漏检概率降低；T 值选择过大，则会导致漏检概率增加而误检率降低。

从数字通信的理论出发，信息隐藏可理解为在一个宽带信道(原始载体信息)上采用扩频通信技术传输一个窄带信号(隐藏信息)。由于隐藏信号的能量较低，它分布到信道中任意特征上的能量是难以检测到的；隐藏信息的检测则可理解为在一个含噪声信道中的弱信号检测问题，如图 4-4 所示。

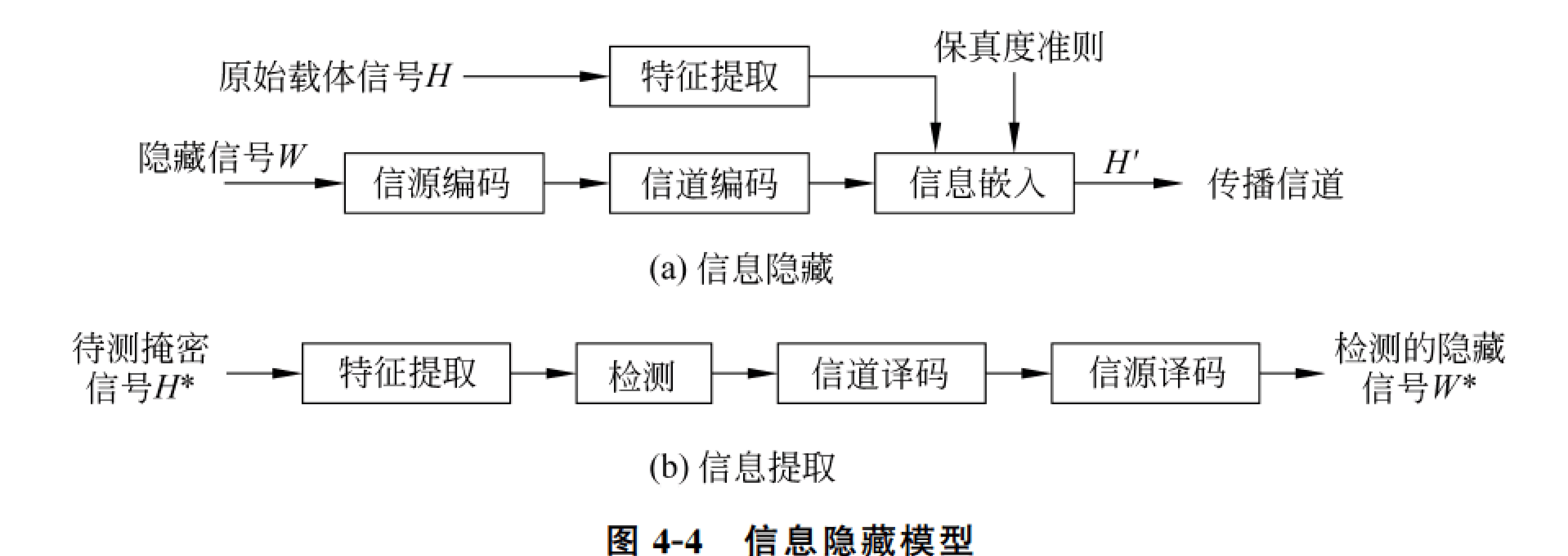


图 4-4 信息隐藏模型

4.2.4 空域信息隐藏技术

空域隐藏技术是指在图像、视频、音频等载体的空间域上进行信息隐藏。通过直接改变宿主媒体的某些像素值(采样值)来嵌入数据。

空域信息隐藏技术无须对原始媒体进行变换，计算简单，效率较高，但由于水印要均

衡不可感知性和稳健性,因而可选择的属性范围较小。此外,难以抵抗常见信号处理的攻击及噪声干扰的影响,鲁棒性较差。

下面介绍两种比较典型的空域信息隐藏方法:基于替换 LSB 的空域信息隐藏方法和 Patchwork 空域信息隐藏方法。

在基于空域的信息隐藏方法中,替换 LSB 位平面的方法是最简单和最经典的一种。LSB(the Least Significant Bits)即最不重要比特位。改变 LSB 主要的考虑是不重要数据的调整对原始图像的视觉效果影响较小。在该方法中,以图像为例,图像部分像素的最低一个或者多个位平面的值被隐藏数据所替换。即载体像素的 LSB 平面根据要隐藏的数据改变为“1”或者不变,以此达到隐藏信息的目的。

令$(f_1, f_2, \cdots, f_n)$为从原始宿主图像中选择出来作为隐藏信息的像素集合,$(b_1, b_2, \cdots, b_n | b_i \in \{0,1\})$为待隐藏的信息,则嵌入过程可表述为:

$$C'(f_i) \leftarrow b_i \tag{4-6}$$

其中算子取载体像素的最低 1 位。

基于替换的 LSB 的隐藏方法具有如下特点:

(1) 具有较大的信息隐藏容量,隐藏容量信息可以达到 1~3 比特/像素。

(2) 计算简单。

(3) 掩密图像失真小。

(4) 隐藏数据的鲁棒性较差。

例 4-1 设待隐藏信息为 1001,取灰度图像的 4 个像素值(0~255 整数)的最低位进行隐藏,如表 4-1 所示。

表 4-1 例 4-1 用表

隐藏前 8 位灰度值	二进制表示	隐藏后二进制	隐藏后 8 位灰度值
34	00100010	00100011	35
180	10110100	10110100	180
255	11111111	11111110	254
2	00000010	00000011	3

Patchwork 算法(亦称为拼凑方法)只是试图回答是否有水印存在,因而实际隐藏的只是 1 比特信息。Patchwork 算法的一般步骤如下:

(1) 用一个密钥初始化一个伪随机数发生器。

(2) 根据伪随机数发生器的输出,随机选择 n 个像素对,其灰度值为(a_i, b_i)。

(3) 令 $a_i^w = a_i + 1, b_i^w = b_i - 1$,完成信息的嵌入。这样整个图像的平均亮度保持不变。检测时,令

$$s = \sum_{i=0}^{N-1} (a_i^w - b_i^w) \tag{4-7}$$

如果 $s \approx 2n$,则判定存在隐藏信息,否则 s 的值应该接近 0。

这种方法是基于如下统计假设的,即原始载体中随机选择的 N 对像素是独立分布

的,也就是满足

$$E[s] = \sum_{i=0}^{N-1}(E[a_i] - E[b_i]) \tag{4-8}$$

因此,只有知道嵌入位置的人才能得到 $s \approx 2n$;否则只能得到 $s \approx 0$。

实际上,该嵌入方法只嵌入了 1 比特信息。通过适当调整参数,拼凑方法对 JPEG 压缩、FIR(Finite Impulse Response)滤波以及图像裁剪有一定的抵抗力,但该方法嵌入的信息量有限。为了嵌入更多的隐藏信息,可以将图像分块,然后对每一个图像块进行嵌入操作。

Patchwork 算法基于改变图像数据的统计特性。该算法首先随机选取 N 对像素点,然后通过增加像素对中一个点的亮度值,而相应降低另一个点亮度值的方法来隐藏信息。为增加水印的鲁棒性,还把像素对扩展为小块的像素区域,通过增加一个区域中的所有像素点的亮度值,从而相应地减少对应区域中所有像素点亮度值的方法来隐藏信息。但该算法嵌入码率低,且对共谋攻击的抵抗能力弱。

4.2.5 变换域信息隐藏技术

变换域隐藏技术中,信息隐藏过程是在变换域中进行的。借助信号进行正交变换后能量重新分布的特点,在变换域中进行信息隐藏,可以较好地解决不可感知性和稳健性的矛盾。因而,基于变换域的方法在信息隐藏研究中占有主要地位。

信息隐藏中的正交变换可以理解为将信号按频谱进行分解,每个分量的值代表信号在此频率上的能量;反变换则是一个对各频率分量进行加权和的合成过程。通常,信号的主要能量集中在低频部分,因而变换域低频系数的值普遍较大,而高频系数则表示信号的突变成分,其系数值相对较小。

信息隐藏中常用的变换有:离散傅里叶变换(Discrete Fourier Transform,DFT),离散余弦变换(Discrete Cosine Transform,DCT)、离散小波变换(Discrete Wavelet Transform,DWT)和 RST 变换。此外,近年还出现 Bandelet 变换、Curvelet 变换等。

1. 离散余弦变换

在通常使用的图像压缩标准,如 JPEG、MPEG-2 等标准中,采用的变换都是 DCT,因此基于 DCT 变换的数字水印技术是切实可行的。基于 DCT 的数字水印算法首先从载体中获取特征进行二维离散余弦变换,然后选择适当的系数将水印嵌入,最后进行二维离散余弦反变换得到加入水印的图像。选择什么样频段的系数是一个很有争议的问题,有人主张将水印加入高频段,这样不至于使原始图像失真;也有人认为应加入到图像的低频段,以增强水印的鲁棒性。现在更为统一的意见是将水印加入到原始图像的中频段以在信噪比和鲁棒性之间折衷。

2. 离散小波变换

由于小波变换具有多分辨分析特点,能充分反映人类的视觉特性,特别是新的图像压缩标准,如 JPEG2000、MPEG-4 等都采用了基于小波变换的方法,因而在小波变换域

研究水印是极为重要的。基于 DWT 的数字水印算法的基本思想和基于 DCT 的数字水印算法的基本思想是基本一样的。但是由于基于 DWT 的数字水印算法具有多分辨特性,水印的嵌入变得更为灵活。

变换域信息隐藏方法的主要步骤如下:

(1) 应用 DCT、DFT、DWT 等方法将原始宿主信号变换到频域空间。

(2) 在变换域选择 n 个系数以隐藏信息。

(3) 根据一定的规则或者公式修改选择的 n 个变换系数。

(4) 进行反变换以得到掩密载体。

与空域的方法相比,变换域的方法有如下优点:

(1) 变换域中嵌入的信号能量可以较均匀地分布到空域的所有像素上,有利于保证不可见性。

(2) 在变换域,HVS(Human Visual System)/HAS(Human Auditory System)的某些特性可以更方便地结合到嵌入过程中,有利于不可感知性和稳健性能的提高。

(3) 变换域的方法可与国际数据压缩标准兼容,从而便于实现在压缩域内的信息隐藏算法。

变换域方法的主要缺点:一般来说,隐藏信息量比空域方法低;计算量大于空域算法;在正变换/反变换计算过程中,由于数据格式的转换,通常会造成信息的丢失,这将等效于一次轻微的攻击,对于隐藏数据量大的情况下,这是不利的。

4.2.6 其他信息隐藏技术

1. RST 域算法

RST 域信息隐藏法的基本思想是利用 Fourier-Mellin 变换,使得经过旋转、缩放、平移后得到的图像和原图像在 RST 域保持一致,它需要先后经过离散傅里叶变换、Fourier-Mellin 变换、DFT,形成的变换域称为 RST 域,然后将水印信号加入 RST 域,最后采用相反的过程先后进行 IDFT、Fourier-Mellin 逆变换、IDFT 得到隐藏信息后的图像。该方法的优点是具有很强的抗几何变换能力,缺点是抵抗有损压缩、低通滤波等信号处理方法的稳健性不够。

2. 压缩域算法

基于 JPEG、MPEG 标准的压缩域信息隐藏系统不仅节省了大量的完全解码和重新编码过程,而且在数字电视广播及 VOD(Video On Demand)中有很大的使用价值。相应地,水印检测与提取也直接在压缩域数据中进行。

3. 网格水印算法

针对计算机图形学中常用的三角形网络模型,提出的网格信息隐藏方案,与多分辨网格处理工具箱集成,不需要建立额外的数据结构和进行额外的复杂计算,直接在网格低频成分中嵌入水印,并且在利用网格处理工具箱进行网格处理时,可以较好地保留隐

藏信息，而且重采样算法简单高效，使简化网格和拓扑结构已改变的网格的水印检测成为可能。

4. 扩频技术

扩频技术的一个重要优点是具有很强的抗干扰性。这一特点对信息隐藏技术特别有用。在数字水印技术中，将原始数据的频域看作通信信道 C，水印看作将通过 C 的信号 S，各种有意、无意的干扰(攻击)看作噪声 N。利用扩频技术原理，将水印分布在许多数据频域系数中，加入每个频域系数的信号能量很小且不可随意检测。然而，水印检测过程知道水印的位置和内容，它能将许多微弱的信号集中起来形成具有较高信噪比的输出值，要破坏水印需要很强的噪声加入所有的频域系数中，但是破坏水印的同时也造成原始数据质量严重下降。

5. 人的生理模型技术

人的生理模型包括人类视觉系统 HVS 和人类听觉系统 HAS。该模型不仅被多媒体数据压缩系统利用，同样可以供信息隐藏技术利用。它的基本思想是利用从模型中导出的 JND(Just Noticeable Difference)描述来确定媒体(图像、声音、视频)的各个部分所能容忍的隐秘信号的最大强度，从而能够避免破坏视觉(听觉)质量，因而这一方法同时具有好的透明性和稳健性。

4.3 数字水印技术

4.3.1 数字水印的框架和分类

数字水印(digital watermarking)是实现数字内容保护的有效方法，已成为多媒体信息安全研究领域的一个热点，也是信息隐藏技术研究领域的重要分支。它通过在原始数据中嵌入秘密信息-水印来证实该数据的所有权。被嵌入的水印可以是一段文字、标识、序列号等。水印通常是不可见或不可察的，它与原始数据(如图像、音频、视频数据等)紧密结合并隐藏其中，成为源数据不可分离的一部分，并可以经历一些不破坏源数据使用价值或商用价值的操作而存活下来。

由于水印信号的嵌入可以视为在强背景下迭加一个弱信号，只要迭加的水印信号强度低于 HVS/HAS 的对比度门限。特别需要注意的是，HVS/HAS 的对比度门限受视觉/听觉系统的空间、时间和频率特性的影响。因此，利用人类的视觉冗余或不敏感性、载体其自身的独特性、数据冗余性、时频的局部特性，把创作者的创作信息和个人标志加入到多媒体数据中，使人们无法从表面上感知加入的信息，只有专用的检测器或计算机软件才可以检测出隐藏的信息，从而达到对数字内容进行保护的目的。

所有嵌入水印的方法都包含这些基本的构造模块，即一个水印嵌入系统和一个水印恢复系统，分别如图 4-5 和图 4-6 所示。其中，密钥可用来加强安全性，避免未授权方恢复和修改水印。

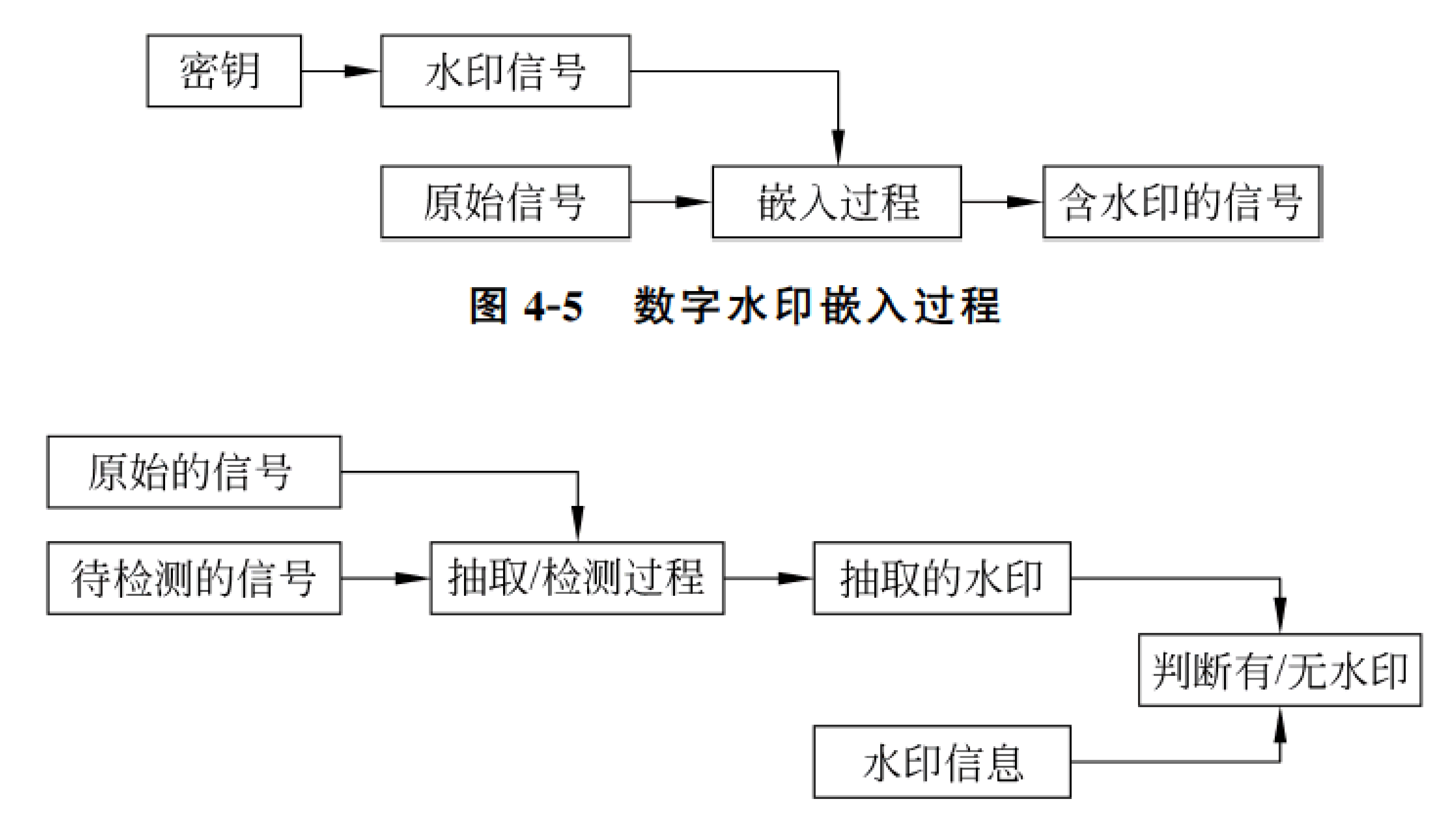


图 4-5 数字水印嵌入过程

图 4-6 数字水印检测/抽取过程

数字水印的分类方法有很多种,分类的出发点不同导致了分类的不同,最常见的分类方法包括以下几类。

1. 按特性划分

按水印的特性可以将数字水印分为鲁棒数字水印(Robust Watermarking)和脆弱数字水印(Fragile Watermarking)两类。鲁棒数字水印主要用于在数字作品中标识著作权信息,利用这种水印技术在多媒体内容的数据中嵌入创建者、所有者的标示信息,或者嵌入购买者的标示(即序列号)。在发生版权纠纷时,创建者或所有者的信息用于标示数据的版权所有者,而序列号用于追踪违反协议而为盗版提供多媒体数据的用户。用于版权保护的数字水印要求有很强的鲁棒性和安全性,除了要求在一般图像处理(如滤波、加噪声、替换、压缩等)中生存外,还需能抵抗一些恶意攻击。

脆弱数字水印与鲁棒水印的要求相反,主要用于完整性保护,这种水印同样是在内容数据中嵌入不可见的信息。当内容发生改变时,这些水印信息会发生相应的改变,从而可以鉴定原始数据是否被篡改。脆弱水印应对一般的信号处理(如滤波、加噪声、替换、压缩等)有较强的免疫能力(鲁棒性),同时又要求有较强的敏感性,既允许一定程度的失真,又要能将失真情况探测出来。脆弱数字水印必须对信号的改动很敏感,人们根据易损水印的状态就可以判断数据是否被篡改过。

2. 按水印所附载的媒体划分

按水印所附载的媒体,我们可以将数字水印划分为图像水印、音频水印、视频水印、文本水印、三维网格模型的网格水印和二维矢量图形水印等。随着数字技术的发展,会有更多种类的数字媒体出现,同时也会产生相应的水印技术。

3. 按检测过程划分

按水印的检测过程可以将数字水印划分为明文水印和盲水印。明文水印在检测过程中需要原始数据,而盲水印的检测只需要密钥,不需要原始数据。一般来说,明文水印

的鲁棒性比较强，但其应用受到存储成本的限制。目前学术界研究的数字水印大多数是盲水印。

4. 按内容划分

按数字水印的内容可以将水印划分为有意义水印和无意义水印。有意义水印是指水印本身也是某个数字图像(如商标图像)或数字音频片段的编码；无意义水印则只对应于一个序列号。有意义水印的优势在于，如果由于受到攻击或其他原因致使解码后的水印破损，人们仍然可以通过视觉观察确认是否有水印。但对于无意义水印来说，如果解码后的水印序列有若干码元错误，则只能通过统计决策来确定信号中是否含有水印。

5. 按用途划分

不同的应用需求造就了不同的水印技术。按水印的用途，我们可以将数字水印划分为票证防伪水印、版权保护水印、篡改提示水印和隐蔽标识水印。

票证防伪水印是一类比较特殊的水印，主要用于打印票据和电子票据、各种证件的防伪。一般来说，伪币的制造者不可能对票据图像进行过多的修改，所以，诸如尺度变换等信号编辑操作是不用考虑的。但另一方面，人们必须考虑票据破损、图案模糊等情形，而且考虑到快速检测的要求，用于票证防伪的数字水印算法不能太复杂。

版权标识水印是目前研究最多的一类数字水印。数字作品既是商品又是知识作品，这种双重性决定了版权标识水印主要强调隐蔽性和鲁棒性，而对数据量的要求相对较小。

篡改提示水印是一种脆弱水印，其目的是标识原文件信号的完整性和真实性。

隐蔽标识水印的目的是将保密数据的重要标注隐藏起来，限制非法用户对保密数据的使用。

6. 按水印隐藏的位置划分

按数字水印的隐藏位置，我们可以将其划分为时(空)域数字水印、频域数字水印、时/频域数字水印和时间/尺度域数字水印。

时(空)域数字水印是直接在信号空间上叠加水印信息，而频域数字水印、时/频域数字水印和时间/尺度域数字水印则分别是在DCT变换域、时/频变换域和小波变换域上隐藏水印。

随着数字水印技术的发展，各种水印算法层出不穷，水印的隐藏位置也不再局限于上述四种。应该说，只要构成一种信号变换，就有可能在其变换空间上隐藏水印。

4.3.2 数字水印的评价指标

(1) 安全性：数字水印的信息应是安全的，难以篡改或伪造，同时，应当有较低的误检测率，当原内容发生变化时，数字水印应当发生变化，从而可以检测原始数据的变更；当然数字水印同样对重复添加有较强的抵抗性。

(2) 隐蔽性：数字水印应是不可知觉的，而且应不影响被保护数据的正常使用；不会

降质；衡量隐蔽性的客观标准有均方误差(Mean-Square Error，MSE)和信噪比(Signal-to-Noise Radio，SNR)。

设原始载体和掩密载体分别用 $H(x,y)$ 和 $H'(x,y)$ 表示，其中 $0\leqslant x\leqslant M-1$，$0\leqslant y\leqslant N-1$，则掩密载体相对于原始载体造成的均方误差 MSE 定义为：

$$\mathrm{MSE}=\frac{1}{MN}\sum_{x=0}^{M-1}\sum_{y=0}^{N-1}\mid H'(x,y)-H(x,y)\mid^2 \tag{4-9}$$

与此同时，水印嵌入载体也可视为在载体中引入了一定的噪声，即：

$$H(x,y)=H(x,y)+e(x,y) \tag{4-10}$$

由此可以定义掩密载体和原始载体的信噪比为：

$$\mathrm{SNR}=\sum_{x=0}^{M-1}\sum_{y=0}^{N-1}H'^2(x,y)\Big/\sum_{x=0}^{M-1}\sum_{y=0}^{N-1}e^2(x,y) \tag{4-11}$$

在此基础上，可以定义出峰值信噪比 PSNR(Peak Signal-to-Noise Ratio)：

$$\mathrm{PSNR}=10\cdot\log\frac{255^2}{\mathrm{MSE}} \tag{4-12}$$

(3) 鲁棒性：是指在经历多种有意或无意的信号处理操作后，数字水印仍能保持部分完整性并能被准确鉴别。可能的信号处理过程包括信道噪声、滤波、数/模与模/数转换、重采样、剪切、位移、尺度变化以及有损压缩编码等。对于脆弱水印，它主要用于完整性保护，这种水印同样是在数字内容中嵌入不可见的信息，当数字内容发生改变时，这些水印信息会发生相应的改变，从而可以鉴定原始数据是否被篡改。

(4) 水印容量：嵌入的水印信息必须足以表示数字内容的创建者或所有者的标志信息，或购买者的序列号，这样有利于解决版权纠纷，保护数字产权合法拥有者的利益。尤其是隐蔽通信领域的特殊性，对水印的容量需求较大。

4.3.3 数字水印的攻击方法

水印攻击与密码攻击一样，包括主动攻击和被动攻击。主动攻击的目的并不是破解数字水印，而是篡改或破坏水印，使合法用户也不能读取水印信息。而被动攻击则试图破解数字水印算法。相比之下，被动攻击的难度要大得多，但一旦成功，则所有经该水印算法加密的数据全都失去了安全性。

主动攻击的危害虽然不如被动攻击的危害大，但其攻击方法往往十分简单，易于广泛传播。无论是密码学还是数字水印，主动攻击都是一个令人头疼的问题。对于数字水印来说，绝大多数攻击属于主动攻击。

值得一提的是，主动攻击并不等于肆意破坏。以版权保护水印为例，如果将嵌入了水印的数字艺术品弄得面目全非，对攻击者也没有好处，因为遭受破坏的艺术品是无法销售的。对于票据防伪水印来说，过度损害数据的质量是没有意义的。真正的主动水印攻击应该是在不过多影响数据质量的前提下，除去数字水印。

1. 解释攻击及对策

解释攻击也称 IBM 攻击或二次水印攻击或水印的死锁，这一攻击是由 IBM 公司

Watson 中心的研究人员首先提出的。因为在一些水印方案中,可能存在对检测出的水印的多个解释。在解释攻击中,载体特征或许被改变或许不被改变。此类攻击往往要求对所攻击的特定的水印算法进行深入彻底的分析。

目前,由解释攻击所引起的无法仲裁的版权纠纷的解决方案主要有四种:第一种方法是引入时戳机制,从而确定两个水印被嵌入的先后顺序;第二种方法是作者在注册水印序列的同时对原始作品加以注册,以便于增加对原始图像的检测;第三种方法是利用单向水印方案消除水印嵌入过程中的可逆性;第四种方法是利用双水印和盲检测技术,杜绝伪造原始图像的可能性。

2. 信号处理攻击及其对策

常见的信号处理攻击法包括无恶意的和常用的一些信号处理方法。现实应用中,会经常对载体采取各种处理以适应不同的要求。以数字图像为例,信号处理攻击法也包括通过加上噪声而有意修改图像以降低图像水印的强度,我们用强度这一术语来衡量嵌入水印信号的幅度相对于嵌入的数据幅度,类似于通信技术中的调制系数这一概念。

解决信号处理攻击的对策有:在人类视觉特性决定的最大容许范围内,增加嵌入的力度;或者采用冗余嵌入技术。两种方法都会增加水印的强度,从而抵抗主动攻击。如以图像为例,把原图分解成相同的几幅小图,在每幅小图上用同样的算法嵌入同一幅水印图像,能有效地增加水印的鲁棒性,提高水印对信号处理攻击的抵抗能力。如果从安全的角度考虑,可以将原图分解成随机大小的小图,在每幅小图上用同样的或不同的算法,嵌入同样的或不同的水印信息,这更能增加水印的对信号处理攻击的抵抗。但同时增加了嵌入信息的数量,在一定程度上影响了图像的质量,也增加了水印检测的难度。

3. 分析攻击及对策

分析攻击法包括在水印的嵌入和检测阶段采用特殊方法来擦除或减弱载体中的水印。这类攻击往往是利用了特定的水印方案中的弱点,在许多例子中,它证明了分析攻击已经成为可能。共谋(collusion attack)或多重文档攻击(multi-document attack)就是这类攻击。以图像为例,共谋攻击用同一图像嵌入了不同水印后的不同版本组合而产生一个新的"嵌入水印"图像,从而减弱水印的强度。

为了防止分析攻击,应该限制提供的水印化数字作品的数量。另外,在水印信号设计中使用随机密钥进行加密也可以有效增加消除攻击的计算复杂度,导致消除攻击不可实现,也可采用图像与水印相关的水印算法。现在提出的许多算法,在水印信号的嵌入位置选择上,基本都采用了随机或伪随机的机制,加强了水印对分析攻击的抵抗能力。

4. 表达攻击及对策

表达攻击有别于其他攻击之处在于它并不需要除去数字内容中嵌入的水印,它是通过操纵内容从而使水印检测器无法检测到水印的存在。实际上在表达攻击中并未改变任何载体的任何信息。

因为大多数水印提取算法需要知道嵌入水印的确切位置,所以表达攻击很难防御。

以图像为例，目前有效的对策是在嵌入水印的同时嵌入水印参照物。那么在提取过程中，先根据水印参照物的变化获得表达攻击的变换步骤，然后应用反转变换获得水印的完整恢复。第二种对策是，使用与图像相关的脆弱水印。当图像被分割时，脆弱水印能报告图像的失真情况。当易损水印不可被检测时，图像的质量也应降低到不可接受的程度。对抗表达攻击的另一个途径是：数字水印在编码时一定存在冗余数据，而冗余数据过多又会影响水印的信息量。最有效地抵抗对策是水印提取算法中，对嵌入水印的位置采用相对的位移地址，而不是采用绝对的存储位置。

4.3.4 版权保护数字水印技术

数字水印技术之所以在近几年中以惊人的速度发展，除了军事、安全方面的原因外，最主要的原动力就是数字作品版权保护的需要。为了解决日趋复杂的版权纠纷问题，现代版权法中出现了所谓“技术措施”和“权利管理信息”两个新概念。技术措施和权利管理信息是版权人采取的权利保护及标示措施，这两个新概念出现在版权法中，是版权保护制度在新技术条件下的发展。数字水印不仅可以作为版权保护的技术措施，而且还提供了对版权管理信息及我国特有的“行政管理信息”的全面支持。

作为一项关系司法认证的技术，尤其是作为标示行政管理信息的手段，数字水印的标准化工作十分重要。从市场经济的角度看，水印技术标准化还意味着相应产品的垄断，谁的技术成为法律认可的标准，谁就理所当然地享有巨大的市场份额。因此，IBM、NEC 等信息产业巨头一直在积极参与有关版权保护水印技术标准的制定工作。

1998 年，美国版权保护技术组织(CPTWG)成立了数据隐藏小组(DHSG)，着手制定版权保护水印的技术标准。在来自各大公司的 7 份技术方案中，DHSG 确定了其中三个作为候选标准。这三个方案是：

(1) IBM 与 NEC 共同制定的技术方案。

(2) Macrovision、Digimarc 和 Philips 联合制定的方案。

(3) Hitachi、Pioneer 和 Sony 共同制定的方案。

虽然 DHSG 进行了大量的技术调研，但它并没有制定技术标准的权利，最终决定数字水印标准的是美国版权保护顾问委员会(CPAC)。IBM、HP、Apple、Microsoft、Intel、Zoran、ATI Tech.、Mediamatics 和 STMicroelectronics 等多家知名企业都是该委员会的会员。

尽管至今还没有形成数字水印的最终技术标准，但 DHSG 已经明确了用于版权保护的数字水印必须满足的一些基本条件，包括：

(1) 隐藏于数字作品中且不可感知。

(2) 可以被专用的数字电路识别。

(3) 不必获取完整数据，仅从数据流中即可检测到数字水印。

(4) 可以标记“未曾复制”、“只可复制一次”和“不能再复制”等复制信息。

(5) 漏检概率低。

(6) 水印内容(字段)的设计必须合理。

(7) 必须使用成熟的技术嵌入或检测水印。

在我国,知识产权问题是一个敏感的话题,只有深入开展数字水印技术的研究,尽快制定我国的版权保护水印标准,才能使我们在未来可能的国际知识产权纠纷中取得主动权。

在信息时代,数字作品的销售过程是相当复杂的,其过程可以简化为如图 4-7 所示的过程。创作者可通过销售商经由网络营销系统(如 Internet)面向客户进行销售。然而,在网络传送数字作品过程中,盗版者可以通过信息处理技术复制与原作品完全一样的拷贝,或者盗版者对其进行进一步处理后转卖给其他客户。作为购买者而言,他们不能保证得到真实的原创数字作品;而作为创作者和销售商来说,他们将蒙受巨大的经济损失。

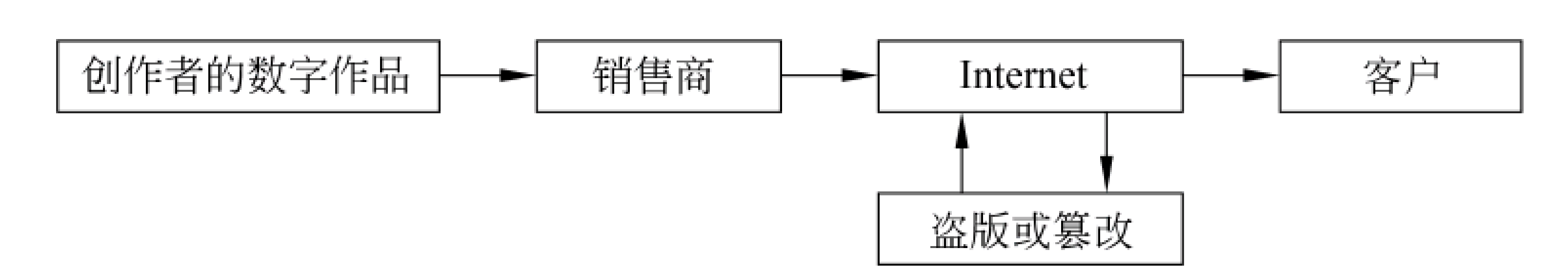


图 4-7　数字产品的销售过程

针对数字作品的销售过程,为保证原创者、销售商到客户的合法权益、防止盗版的产生,基于数字水印的数字作品保护一般模型如图 4-8 所示。

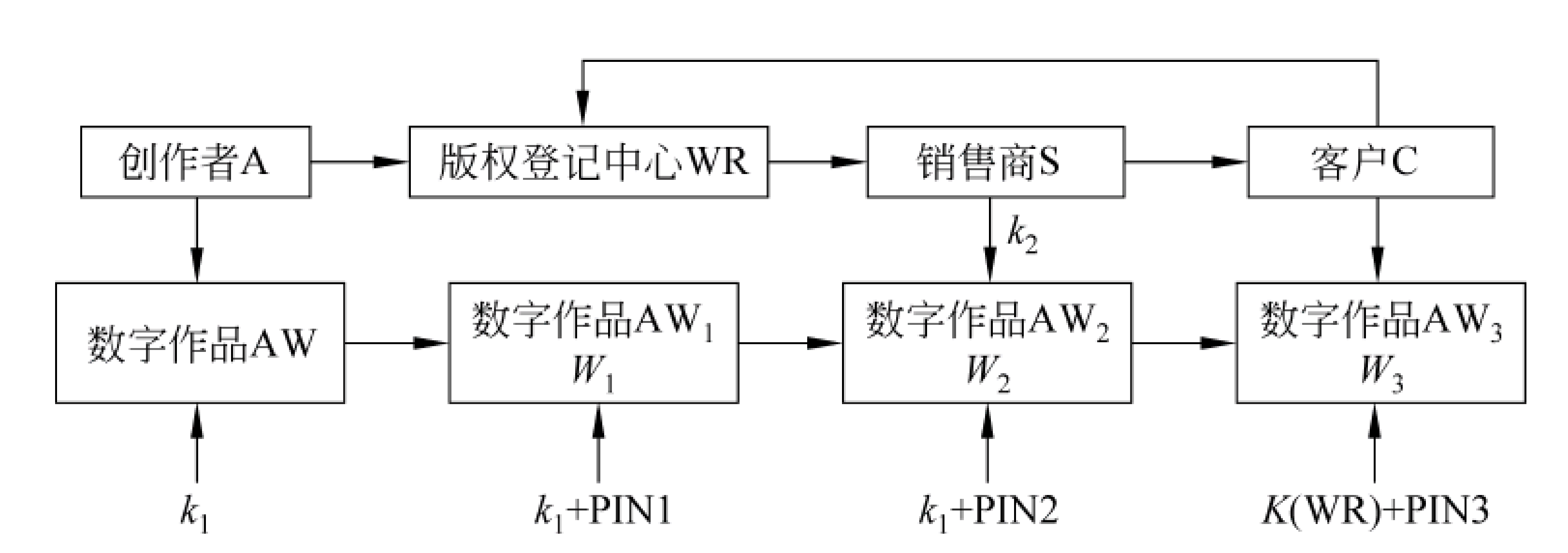


图 4-8　基于数字水印的数字作品保护一般模型

设数字作品的创作者为 A,版权登记认证中心为 WR,A 创作出数字作品后,向 WR 进行作品登记,然后选择一个 A 个人用的私钥 k_1 向期望保护的数字作品 AW 嵌入含有 A 的标志(PIN1)的第一个数字水印 W_1,再将加过水印的数字作品 AW_1 传一份备份给 WR 的数据库中,k_1 由 A 的口令产生,具有唯一性。当 A 决定将其数字作品授权给数字媒体销售商 S 时,让 S 销售其作品的复制品(即拷贝)时,A 需要将 S 的标志(如 PIN2)结合私钥 k_1 对数字作品嵌入第二个数字水印 W_2,以表示对 S 的授权和认可。S 得到加有两个数字水印标志的数字作品,并也可以用 A 的公钥 k_2 验证 A 确实在其作品的拷贝中加入了 S 的标志,即 W_2。S 作为 A 的数字作品销售商,可以应用检测水印的软件,验证第二个水印的内容和第一个水印的内容,但 S 并不感兴趣破坏水印的内容,因为这将破坏他的利益。

授权的 S 将数字作品售给授权用户 C,为证明 C 经过授权,为正版用户,S 用 WR 的私钥 K(WR)和 C 的标志(PIN3)对作品嵌入第三个水印 W_3,并将此信息通知 WR,WR 发给 S 一个证书,给 A 增加一份收益。在此过程中 WR 充当认证中心的作用。

4.3.5 内容认证数字水印技术

由于内容认证数字水印要检测出篡改位置并进行定位,因此,通常嵌入的水印信息是与原始载体内容相关的信息(但也可以是不相关的)。其嵌入过程是:首先对原始载体进行特征提取并以此来构造水印信息,再将水印信息嵌入到原始载体中就得到嵌入水印后的受保护数字内容,其水印嵌入过程如图 4-9 所示。

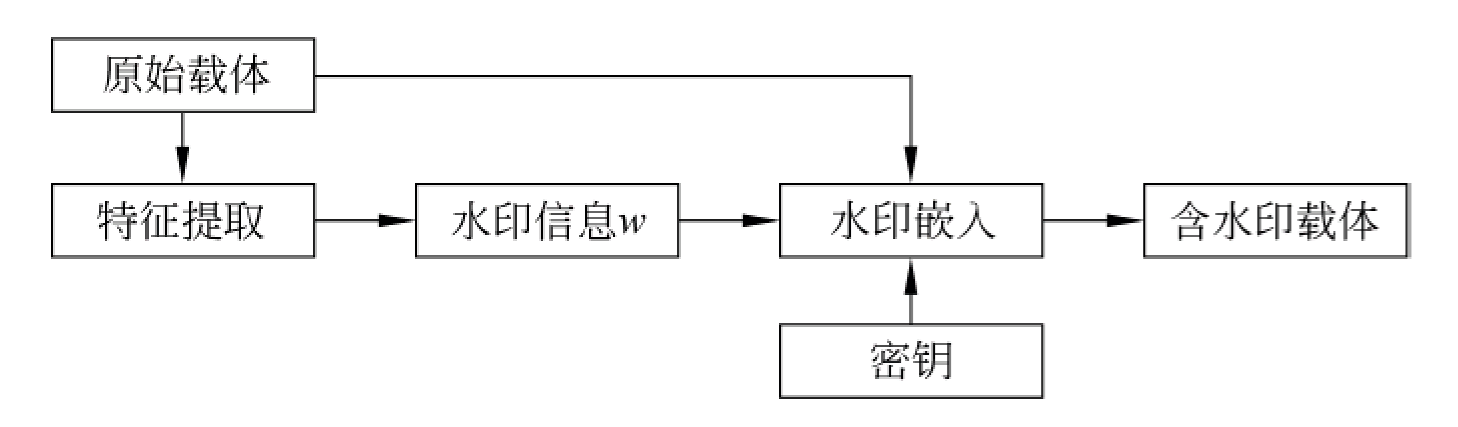


图 4-9 内容认证数字水印的嵌入过程

图像内容认证过程如图 4-10 所示。对数字内容进行认证时,根据密钥提取出受保护图像中的水印信息。然后将提取出来的水印信息与原始水印信息相比较,若二者一致,则图像未被更改;若二者不一致,则认为图像已被更改,并给出有关图像改动的详细信息。如果嵌入的水印信息是与原始图像内容相关的信息,并确保水印的嵌入不会改变图像的这些内容特征,则图像认证时,只需将提取出的水印信息与被测图像的内容特征进行比较。目前各种水印认证算法主要在水印信息的生成和嵌入两个过程存在一定的差异。

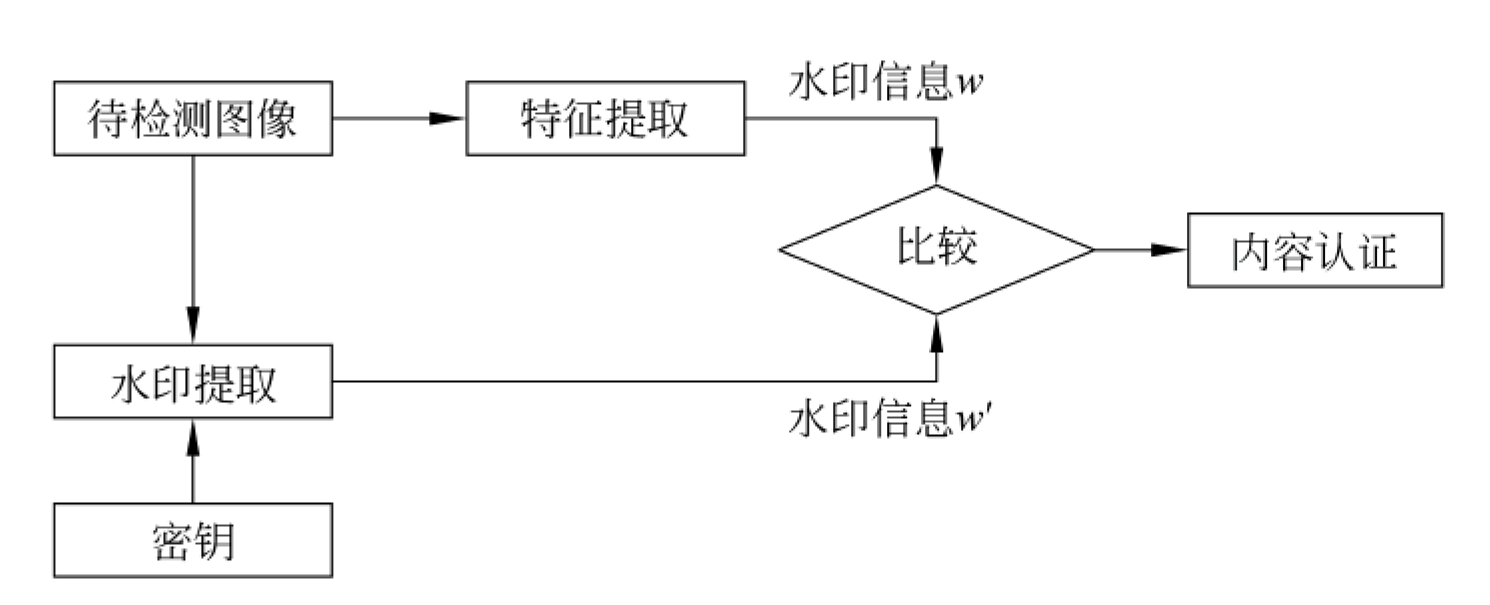


图 4-10 图像内容认证过程

4.3.6 可逆水印技术

可逆数字水印(reversible watermark)技术属于数字水印技术的一个分支,目前大多数数字水印的方法在提取出所嵌入的秘密信息后,原宿主信息不能无损恢复,属于有损数字水印技术。但是在一些要求较高的场合,如医学诊断、军事图像、遥感图像处理和法律认证及证据等领域,则往往需要精确地恢复原载体。例如,在法律上,为了保证法律的绝对权威性和强制性,作为证据的数字载体对象的丝毫改变都将可能影响法律的公正性。在医学成像领域中,由于医生对图像的错误认识会产生的潜在危险,使得即使很小的图像修改都是不允许的。在军事上,军事图像分析专家可能需要在特殊的观察条件下

检查图像，这时，对于原始载体数据的微小的破坏都可能引起严重的后果。在这些特殊的应用场合，可逆数字水印技术作为解决这些问题的一个有效途径而得到了重视和研究。为此很多学者已开展了大量的关于可逆数字水印（也叫无损数字水印）方面的研究。可逆数字水印要求通过嵌入信息后的载体，不仅可以提取该载体中隐藏的秘密信息，而且还可以实现原始载体的完全恢复重构。

1. 可逆数字水印方法分类

常见的可逆数字水印方法可分为基于无损压缩的可逆水印方法、基于差值扩展的可逆数字水印方法和基于直方图修改的可逆数字水印方法等。

1）基于无损压缩的可逆数字水印方法

基于载体图像的最低有效位（Least Significant Bit，LSB）无损压缩后和水印信息一起用 LSB 替换嵌入到载体图像中，在此基础上随后提出了一种 G-LSB 方法，该方法首先将图像量化，然后将量化后的差值用 CALIC 无损压缩后和秘密信息一起嵌入到载体图像中，使嵌入量大大提高。一种基于哈尔小波的可逆数字水印方法，该方法通过将变换域中高频数据的整数部分进行无损压缩后，和秘密信息一起经 2 位的 LSB 替换嵌入到其高频整数部分，在该方法中，由于高频部分大部分整数部分一般较小，基于哈尔小波的可逆数字水印方法的嵌入率较前几种方法有所提高。这一类方法的嵌入量都和压缩率紧密相关，嵌入率较低。

2）基于差值扩展（difference expansion）的可逆数字水印方法。

采用整数小波变换和差值扩展的方法，将一比特数字水印嵌入在两个相邻的像素点中，所以嵌入一次的嵌入率约为 0.5bpp。基于整数小波阈值的可逆数字水印方法将秘密信息嵌入到小波变换后的根据阈值选取的高频子带系数中，并采用直方图调整的方法防止数据的溢出。对差值扩展技术进行改进，通过低通滤波单元预测要扩展的位置，所以该方法对数据修改量小，同时嵌入量也不高。通过对像素点和其预测值的差值提出一种基于预测误差扩展的可逆信息方法，取得了较大的嵌入率。随后有人提出一种基于中心差值扩展（centralized diference expansion）的方法，该方法根据大小和复杂度将分块后图像块分为四类，然后根据不同的情况进行自适应嵌入，该方法在具有较好的隐藏视觉效果的同时具有较高的嵌入率。用线性预测误差扩展和改进的嵌入/提取算法，在图像无损编码压缩过程中嵌入大数据量的秘密信息。选择图像的边缘和纹理进行差值扩展嵌入，并结合压缩性非常高的溢出位置图，实现了高视觉质量、大嵌入容量的可逆数据隐藏算法。

3）基于直方图修改的可逆数字水印方法

在这种方法中，首先找到直方图的峰值点和零点，然后通过直方图修改将秘密信息嵌入在具有峰值点灰度值的像素点中。该方法具有很高的峰值信噪比（PSNR），高于 48.13dB，但一般情况下嵌入量较低。在一种多层的基于差值图像直方图修改的可逆数字水印方法中，该方法首先将图像分块，然后在各个子图像块中采用直方图修改的方法进行隐藏，该方法具有较高的嵌入率，然而其密钥（各个子图像块的峰值点）的数据量巨大，如果去除密钥信息量，该方法几乎没有什么优势可言。在一种基于预测和差值直方

图的可逆数字水印方法中,该方法首先将图像分块,然后计算图像块中各个像素值和块中间像素点的差值,并计算其差值的直方图,采用直方图修改方法将秘密信息嵌入在该差值中。

第三种方法需要附加信息的信息量小,而且实现也比较简单,同样引起了研究者的广泛关注。此外,还出现了一些具有特殊作用的可逆数字水印方法,如对JPEG压缩具有一定鲁棒性的可逆数字水印方法。基于特定图像格式的可逆数字水印方法,如基于VQ压缩的可逆数字水印方法,基于JPEG压缩的可逆数字水印方法。

可逆数字水印相对于有损数字水印,具有如下优势,一是可实现原始载体的无损恢复,因此可逆数字水印在医学军事图像、遥感图像处理和法律认证及证据等领域具有很强的应用价值。二是多层数字水印,有损数字水印只能进行一层数字水印,因为如果将有损载体进行二次隐藏后,会破坏第一层隐藏的数据,而可逆数字水印由于其可完全始载体数据,所以可实现多层可逆数字水印。

2. 可逆数字水印的评价指标

可逆数字水印主要有两方面的性能评价:畸变程度和嵌入率(隐藏量)。

1)畸变程度

在数字水印方法中,秘密信息的嵌入会不可避免地使秘密信息嵌入后的载体产生一定程度的畸变。评价秘密信息嵌入后的不可见性有两种方式,一种是主观评价法,另一个是客观评价法。主观评价法是将人对载体的感觉分为几个等级,然后综合几十个人对载体的直接感觉来综合评价载体的质量,但主观评价法的评价结果容易受到评价者主观因素的影响;因此对于不可见性的评价需要客观的评价方法。常用的方法主要有均方差和峰值信噪比。

峰值信噪比是一个表示信号的最大可能功率与影响它表示精度的破坏性噪声功率的比值。峰值信噪比越大,说明数字水印的不可见性越好,峰值信噪比越大,说明载体的质量降低得越少。但这种评价没有从根本上反映出载体处理前后在视觉上的变化情况。由于人眼的视觉特性受到外界条件的影响,不同的光照度、不同的背景都会影响人眼的感知,而且人眼对图像是一个整体的感知,周围像素点会影响人眼对该像素点的视觉效果。总之,峰值信噪比与人眼的视觉并没有必然的相关性。由于人眼是数字水印最直接的审判者,所以用峰值信噪比来评价隐藏后的载体效果不是十分合适,但目前并没有其他更好的方法可用,通用的方法是采用主观评价和客观评价相结合的方式进行综合评价。

2)嵌入率

在保证被隐藏的秘密信息不可见的情况下,嵌入量也是极其重要的。嵌入量就是嵌入到载体中的秘密信息的比特数。在一般的可逆数字水印方法中,常常会产生附加信息。所谓附加信息是指为了恢复原始载体,需要传送给接收者和秘密信息一起嵌入到载体中的信息。接收者通过附加信息将秘密信息提取出来并恢复原始载体。嵌入量可以分为实际嵌入量和有效嵌入量。实际嵌入量即数字水印方法本身可嵌入的比特数,而有效嵌入量是指实际可隐藏的嵌入量再减去附加的信息量。如果嵌入量只考虑实际嵌入

量而不考虑一些要传送给接收者的附加信息是不合理的。

常用来衡量有效嵌入量的指标是嵌入率 ER(Embedding Rate),其表达式为:

$$ER = \frac{Num_{sec} - Num_{extra}}{Num_{feature}}(bit/feature,bpf) \tag{4-13}$$

式中,Num_{sec}表示嵌入的秘密信息的比特数;Num_{extra}表示附加信息的比特数;$Num_{feature}$表示图片载体的特征点个数。这里的 feature 在不同的载体对象中有不同的表示,例如在点阵图像中为像素(Pixel),在矢量图像中为顶点(Vertex)。嵌入率可以较直观地反映一种方案嵌入能力的大小。

3. 基于差分扩展的可逆水印技术

1) 差值扩展的基本原理

基于差值扩展的可逆数字水印方法最早是在 2003 年被提出来的,秘密信息被嵌入在整数小波域的高频系数(差值)上,为了防止逆变换后数据溢出问题的出现,需要选取合适的高频系数,并且采用差值扩展的方式实现秘密信息的嵌入。此外,为了实现原始载体图像的无损恢复,嵌入的位置图记录了秘密信息嵌入的位置。以图像为例,基于差值扩展的可逆数字水印方法的基本步骤是将图像分成像素点对(x,y),其中 $x,y\in Z$,$0\leqslant x,y\leqslant 255$。根据下式定义其整数均值 l 和差值 h:

$$\begin{cases} l = \text{floor}\left(\dfrac{x+y}{2}\right) \\ h = x - y \end{cases} \tag{4-14}$$

其中,floor()表示向下取整,式(4-13)的逆变换为:

$$\begin{cases} x = l + \text{floor}\left(\dfrac{h+1}{2}\right) \\ y = l - \text{floor}\left(\dfrac{h}{2}\right) \end{cases} \tag{4-15}$$

式(4-14)的整数变换也叫做整数 Haar 小波变换(Integer Haar Wavelet Transform, IHWT)。该整数变换在整数对(x,y)和(l,h)之间建立了一对一的映射。

将秘密比特信息 m 以差值扩展的方法嵌入到差值 h 中:

$$h' = h \times 2 + m \tag{4-16}$$

将式(4-16)得到的 h'代入式(4-14)中,得到新的图像像素对,形成嵌入秘密信息后的图像。

为了保证可逆隐藏后的数据没有溢出,设载体图像为 8 位的灰度图像,可根据嵌入秘密信息后的恢复值不能超过[0,255]范围求出,即:

$$\begin{cases} 0 \leqslant l + \text{floor}\left(\dfrac{h+1}{2}\right) \leqslant 255 \\ 0 \leqslant l - \text{floor}\left(\dfrac{h}{2}\right) \leqslant 255 \end{cases} \tag{4-17}$$

根据上式,可推出其可逆数字水印不溢出的条件:

$$\begin{cases} |h| \leqslant 2(255 - l), & 128 \leqslant l \leqslant 255 \\ |h| \leqslant 2l + 1, & 0 \leqslant l \leqslant 127 \end{cases} \tag{4-18}$$

由于嵌入后的差值为 $h'=h\times2+m$，将其代入式(4-14)可得某像素点对进行可逆数字水印的条件：

$$|2\times h+b|\leqslant\min(2(255-l),2l+1) \tag{4-19}$$

在差值扩展的可逆数字水印方法中，将差值图像分为可扩展差值、可改变而不可扩展差值和不可改变差值。选择小幅值的差值作为嵌入位置，这样可减小图像的畸变。根据负载的大小预先设定一个阈值，选择小于该阈值的可扩展差值为候选位置，并用一个尺寸与高频图像一样的二值矩阵作为嵌入位置图，用来标识可扩展和非可扩展的差值位置。最后，压缩后的位置图比特流需要和秘密信息一起嵌入到差值图像中。嵌入完成后，进行小波逆变换，可得到嵌入信息后的图像。

2）差值扩展方法的优缺点

总结差值扩展的可逆数字水印方法，具有以下优点：

① 算法简单，容易实现。只要对原差值乘以 2，再进行 LSB 嵌入即可实现嵌入过程。

② 差值数据的直方图自动向两边漂移。当 h 大于等于 0 时(b 为 0 或 1)，可以证明 h'必大于等于 0；当 h 小于 0 时，h'也小于 0。图像的像素点差值向两边漂移时，保证每对像素点的值相对变化均匀。

③ 满足式(4-19)的像素点对的个数决定其隐藏量。对于一般图像，相邻像素点具有高相关性，其像素点对的差值一般较小，不难满足式(4-19)，其隐藏量大致可认为是该图像分成的像素点对的个数。

由于基于差值扩展的可逆数字水印具有算法简单，容易实现，且隐藏量大，所以受到了广泛的关注。然而，它仍然还存在如下缺点：

① 附加信息过多。在基于差值扩展的可逆数字水印方案中，采用式(4-16)将其直方图向两边漂移，会产生较多的附加信息，而且情况比较复杂。在 Tian 的方案(见文献[23])中，将图像分成了几种状态，将图像可嵌入与不可嵌入信息的部分用一个单比特标识图(bitmap)来标明，再将其无损压缩后，作为附加信息传输。

② 盲目地扩展。将差值都乘以 2 后，相当于将所有差值的直方图都进行了移动，将满足条件的所有像素点对形成的差值都进行隐藏，然而对于出现较少个数的差值也留出空位隐藏信息，这是没有必要的，因为有些差值可隐藏的数据量很小，小到仅仅有几比特，并且个数较少的差值一般较大，秘密信息扩展嵌入后对数据的改变量也会较大。

4. 基于直方图修改的可逆数字水印方法

这里，以图像为例，介绍一种基于直方图修改的可逆水印方法的基本原理。水印嵌入的步骤如下：

① 首先计算图像的直方图，并找到其中的零点，记为 z。零点指的是图像中没有任何一点的灰度值等于 z。然后找到直方图的峰值点，记为 p。峰值点指的是图像中具有像素点最多的灰度值。为了方便叙述，不妨假设 $p<z$。

② 由上到下、由左到右扫描图像中的各个像素点，各个像素点的灰度值用 v_{ij} 表示，当 $v_{ij}<p$ 或 $v_{ij}>z$ 时，像素点的值保持不变，即 $v'_{ij}=v_{ij}$；当 $p<v_{ij}<z$ 时，像素点的灰度值加 1，即 $v'_{ij}=v_{ij}+1$。图 4-11 给出了图像 Lena 的直方图的峰值点及漂移后的直方图。

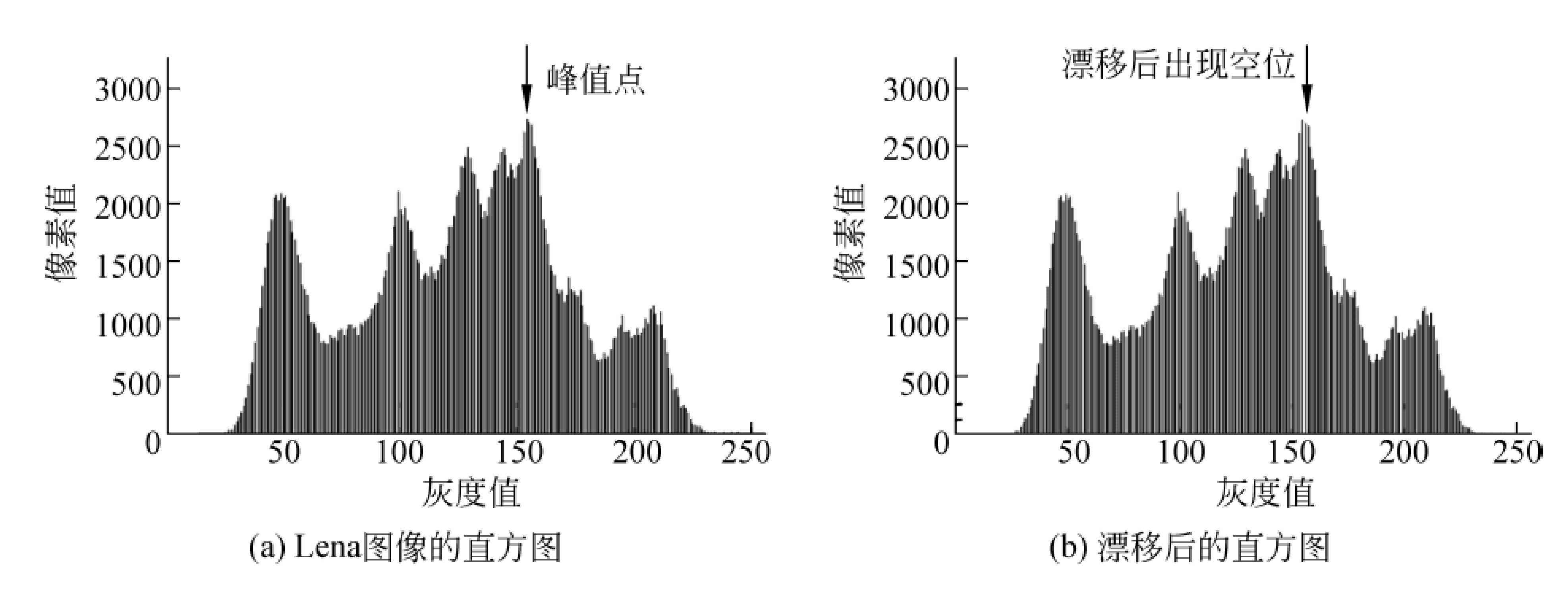


图 4-11　原始直方图和漂移后的直方图

③ 图像中灰度值等于峰值点(即 $v_{ij}=p$)的像素点，为可嵌入秘密信息的点，将秘密信息转化为二进制流，用 s_k 表示。顺序嵌入信息后得 $v'_{ij}=v_{ij}+s_k$。

④ 得到的由灰度值 v'_{ij} 组成的图像就是嵌入秘密信息后的图像。同时 p、z 以密钥的形式保存。

⑤ 可能会造成一些像素点灰度值变化过大，譬如在图像的边缘或变化不平滑的图像区域，像素点对做差得到的 h 值可能较大，这样，采用式(4-16)扩展后的 h' 的变化也就会较大，单个的点对发生变化较大，因此对原始图像的修改量较大，使 PSNR 大大降低。

该方法的隐藏量为直方图峰值量，并且该方法对载体图像每个像素点的修改量最大为 1。因此，根据峰值信噪比的定义可知：

$$\text{PSNR}\geqslant 10\lg\left[\frac{M\times N\times 255^2}{M\times N\times 1^2}\right]=48.13(\text{dB})$$

即在最坏的情况下，峰值信噪比 PNSR=48.13dB。

秘密信息提取和原始图像的恢复过程如下：

① 读取密钥，得到 p、z 的值。

② 逐行扫描图像，各个像素点的灰度值同样用 v_{ij} 表示。当 $v_{ij}=p$ 时，说明该点为隐藏信息的点，提取信息“0”并保持该点灰度值不变；当 $v_{ij}=p+1$ 时，该点也为隐藏信息的点，提取秘密信息“1”并使该像素点灰度值减 1，即 $v'_{ij}=v_{ij}-1$。

③ 当 $v_{ij}<p$ 或 $v_{ij}>z$ 时，像素点的值保持不变，即 $v'_{ij}=v_{ij}$；当 $p-1<v_{ij}<z$ 时，像素点的灰度值减 1，即 $v'_{ij}=v_{ij}-1$。

④ 得到的由灰度值 v'_{ij} 组成的新图像就是提取秘密信息后的恢复出来的载体图像。

该方法可实现可逆信息隐藏，是一种有效的可逆信息隐藏方法。

基于直方图修改的可逆数字水印方法具有以下优点：

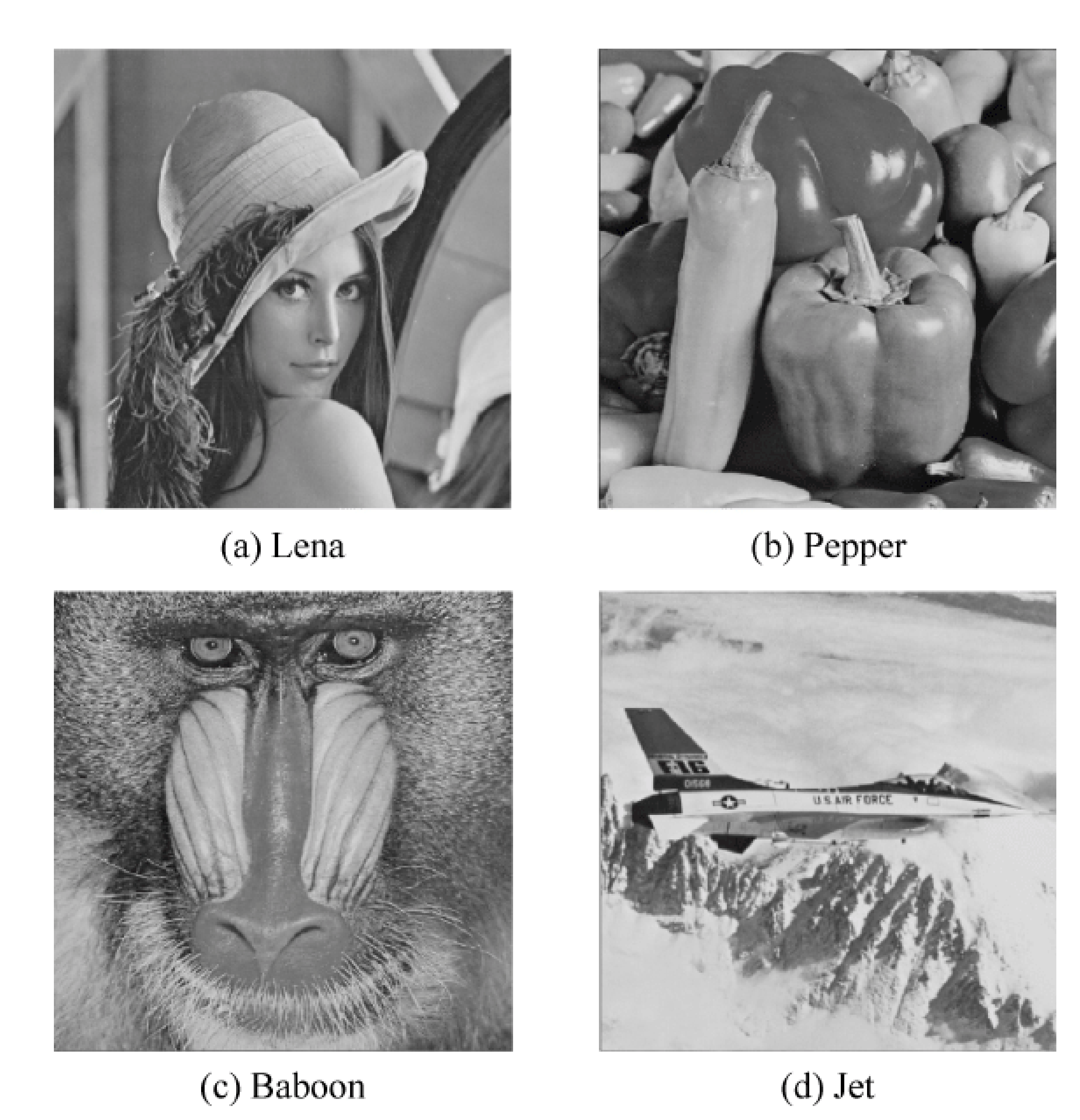
(1) 产生较少的畸变，具有较高的峰值信噪比。如前所述，嵌入一次后其峰值信噪比不低于 48.13dB。

(2) 该方法直接在空域中应用不会产生数据溢出问题。该方法的数据溢出问题较容易解决，因为其溢出的形式单一、容易判断。

(3) 对于某些含有大量相同背景的图像具有较高的嵌入率。这对于一些数字医学图像具有很好的应用效果。

当然,基于直方图修改的可逆数字水印方法也存在缺点:由于图像直方图的直接使用,因此其嵌入率不稳定,对于一般图像嵌入率较低,因为该方法中载体图像的直方图的峰值直接决定其嵌入量。

例 4-2 给定 8 位的大小为 515×512 的载体图像(见图 4-12),表 4-2 给出了其相应的隐藏后的峰值信噪比及嵌入率,由表可见只有当载体图像为 Jet 时,其嵌入率为 0.0317bpp,而对于其他三幅图像,嵌入率都只有 0.0105bpp,嵌入率较低。



(a) Lena (b) Pepper

(c) Baboon (d) Jet

图 4-12 原始图像

表 4-2 直方图修改的隐藏结果

载体图像	Lena	Jet	Pepper	Baboon
PSNR(dB)	53.94	54.48	50.08	50.94
嵌入率(bpp)	0.0105	0.0317	0.0105	0.0105

由以上分析可见,基于直方图修改的可逆信息隐藏方法具有较好的不可见性,但是嵌入率较低。为了使其嵌入率提高,其直方图的峰值点应具有较高的峰值。由于对于大多数图像,其相邻像素点的灰度值具有极大的相关性,因此许多改进方法的思想多集中于此。

差值扩展是把图像分成像素点对,将秘密信息隐藏在该像素点对中。利用该思想,计算像素点对的直方图,其差值的直方图通常具有很高的峰值,而且其值很紧凑,均集中在零值附近,具有很好的聚集性。对大量的图像进行了同样的统计实验,结果表明所得到的差值图像的直方图具有如下共同特点:

(1) 差值直方图的零点较多,且多分布在两端。

(2) 差值直方图的峰值点一般较小。

(3) 差值的集中度增大,具有较高的峰值点,而且其峰值比原始图像的直方图峰值高

得多。

根据以上的特点，可以对原有的直方图修改做如下改进：

(1) 选择无穷远处为其差值直方图的零值点。已知直方图修改法的零值点和峰值点是要以密钥的形式传到接收端。当在一个峰值隐藏时，采取以下方法让尽量少的点发生移动：当在(0,255)区间内的值较少时，可选择正无穷远处为零值点；当在(−255,0)区间内的值较少时，可选择负无穷远处为零值点。由于密钥越少越好，若将零点设为无限远，密钥就不必再包含零点信息，这样可使密钥的长度减少一半。

(2) 可选择两个或多个峰值进行可逆信息隐藏。由于直方图的峰值一般在中间，当利用一个峰值隐藏时，图像中将有一半的像素点发生漂移。当利用两个峰值时，可以选用正负无穷远处的零点，分别利用较大的峰值点和正无穷远处的零点、较小的峰值点和负无穷远处的零点进行两次隐藏，这样每次可嵌入两个最大峰值和的隐藏量，并且保证了每一个像素点的改变量都小于等于 1。当利用两个峰值进行隐藏时，最低的峰值信噪比 PSNR 为 48.13dB。然而，如果选用两个以上的峰值进行可逆信息隐藏时，其最低的峰值信噪比将比 48.13dB 小。

(3) 可以进行多层隐藏。当隐藏数据较多的时候，进行一层隐藏满足不了要求，有以下两种增加隐藏量的方案：一种是多峰值隐藏；一种是多层隐藏。如果每层嵌入时都采用两个峰值进行隐藏，那么进行两层隐藏时，像素点灰度值的改变量为 2，其最低的峰值信噪比变为：

$$\mathrm{PSNR} \geqslant 10\lg\left[\frac{M \times N \times 255^2}{M \times N \times 2^2}\right] = 42.11(\mathrm{dB})$$

同理可计算出当进行 n 层隐藏时，每层选用两个峰值点，并分别以正负无穷远为零值点进行隐藏时的峰值信噪比的最低值(如表 4-3 所示)。

表 4-3　改进的直方图隐藏结果

隐藏层数	1	2	3	4	5	6	7
PSNR(dB)	48.13	42.11	38.58	36.09	34.15	32.57	31.23

表 4-3 比较了直接采用直方图修改和进行差值后采用直方图修改的可逆信息隐藏方法，在选用一个峰值，但载体图像不同时的嵌入率和峰值信噪比。由表 4-4 可见，差值后进行直方图修改的可逆信息隐藏方法(Histogram Modification based on Difference, HMD)的嵌入量远远高于直接采用直方图修改的可逆信息隐藏方法(HM)，两种方法的峰值信噪比基本相同，尤其对于比较平滑的图像。

表 4-4　两种可逆信息隐藏方法的效果比较

载体图像	Lena		Airplane		Sailboat		Baboon	
方法	HM	HMD	HM	HMD	HM	HMD	HM	HMD
嵌入量(比特)	2726	20283	7905	31209	3707	15004	2757	8385
PSNR(dB)	53.70	51.93	50.61	52.55	53.78	51.69	50.61	51.43

5. 变换域中可逆数字水印方法的研究

可逆数字水印要求在提取出秘密信息的同时可无损恢复原载体，这样就要求在数据变换即秘密信息嵌入过程中就需要没有小数舍入，否则很难保证嵌入秘密信息后的载体数据的值都保持为整数。以图像为例，当嵌入信息在频域上做出修改，再乘以逆变换矩阵，就会使隐藏秘密信息后的图像的灰度值出现小数，这样就必须用舍入方式把图像变成整数形式，这样的变换会使数据受损，这些对数据的破坏无法保证提取的秘密信息不受到影响。所以现有的基于变换域的可逆数字水印方法一般都采用整数变换，即变换与逆变换都为整数变换。

1）变换域中可逆数字水印的实现

有些变换可实现可逆数字水印，而有些变换却不能实现。在二维图像中，对图像的变换通常是对图像矩阵左乘、右乘或者既左乘又右乘的变换，这里，以左乘一个矩阵为例进行说明：

$$\boldsymbol{S} = \boldsymbol{T} \cdot \boldsymbol{C}$$

式中，$\boldsymbol{T}$ 为转换矩阵，$\boldsymbol{C}$ 为原始矩阵，$\boldsymbol{S}$ 为变换后的矩阵。

在变换域中采用差值扩展的方法实现可逆数字水印，相当于在图像的高频数据的整数部分乘以 2 后加上秘密信息后作为该数据的新的整数部分。所以矩阵中每一个数据的整数部分的变化可近似为：

$$\Delta = 2 \times h + b - h = h + b$$

在变换域中采用直方图修改方法实现可逆数字水印，直方图向右漂移，对应的原始数据加 1，直方图向左漂移，对应的数据减 1，嵌入信息的点则变化 0 或 1。

上述两种方法有一个共同的特点，即原始数据以 1 为最小单位变化。设变换域内的原始矩阵为 $\boldsymbol{S}$，嵌入秘密信息后的矩阵为 $\boldsymbol{S}'$，嵌入信息的过程就可以看做是在 $\boldsymbol{S}$ 上叠加了一个 $\boldsymbol{D}$ 矩阵（$\boldsymbol{D}(i,j) \in \boldsymbol{Z}$）。因此，无论在空域或变换域，嵌入的数学模型可写为：

$$\boldsymbol{S}' = \boldsymbol{S} + \boldsymbol{D} \tag{4-20}$$

其逆变换为：

$$\boldsymbol{C}' = \boldsymbol{T}^{-1} \cdot \boldsymbol{S}' = \boldsymbol{T}^{-1}(\boldsymbol{S} + \boldsymbol{D}) = \boldsymbol{C} + \boldsymbol{T}^{-1}\boldsymbol{D} \tag{4-21}$$

如果在变换域中利用差值扩展、直方图修改或者压缩整数部分的方法实现可逆数字水印，必须保证 $\boldsymbol{C}'$ 为整数，由式(4-21)可见，只要 $\boldsymbol{T}^{-1}\boldsymbol{D}$ 为整数，就可保证 $\boldsymbol{C}'$ 为整数，由于 $\boldsymbol{D}$ 为附加的整数矩阵，所以只要矩阵 $\boldsymbol{T}^{-1}$ 中所有的元素都为整数，就可以保证 $\boldsymbol{T}^{-1}\boldsymbol{D}$ 为整数。

如果变换矩阵的逆矩阵中的元素都为整数，可采用压缩、直方图修改或者差值扩展的方法在其变换后的数据的整数部分嵌入秘密信息后，再经其逆变换，就可得到不经任何近似的只有整数的图像数据，从而实现可逆数字水印。

2）小波分析中的变换矩阵

对于两个数据 a、b，其 Haar 离散小波变换的过程可表示为：

$$\begin{cases} L = \dfrac{a+b}{2} \\ H = \dfrac{a-b}{2} \end{cases} \tag{4-22}$$

式中，L 表示低频参数，H 表示高频参数。

二维 Haar 离散小波变换是利用一维小波变换的方式先按行变换得到 L 和 H 两个频带，再按列进行变换，得到 LL、LH、HL、HH 四个区域，设有 a、b、c、d 四个数据构成的二维数据，变换结果为图 4-13 所示，其变换可表示为：



图 4-13　Haar 小波变换示意图

$$\begin{cases} \mathrm{LL} = \dfrac{a+b+c+d}{2} \\ \mathrm{LH} = \dfrac{a+b-c-d}{2} \\ \mathrm{HL} = \dfrac{a-b+c-d}{2} \\ \mathrm{HH} = \dfrac{a-b-c+d}{2} \end{cases} \tag{4-23}$$

这是一个线性变换，可写成矩阵形式：

$$\begin{bmatrix} \mathrm{LL} & \mathrm{HL} \\ \mathrm{LH} & \mathrm{HH} \end{bmatrix} = \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ 1/\sqrt{2} & -1/\sqrt{2} \end{bmatrix} \begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ 1/\sqrt{2} & -1/\sqrt{2} \end{bmatrix}^{\mathrm{T}} \tag{4-24}$$

式中，对原始的像素点进行的行和列的变换相同，相当于对原始的矩阵左乘一个 Haar 变换矩阵，再右乘一个 Haar 变换矩阵的转置。通过对变换矩阵求逆，易得到 Haar 逆变换表达式：

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ 1/\sqrt{2} & -1/\sqrt{2} \end{bmatrix} \begin{bmatrix} \mathrm{LL} & \mathrm{HL} \\ \mathrm{LH} & \mathrm{HH} \end{bmatrix} \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ 1/\sqrt{2} & -1/\sqrt{2} \end{bmatrix}^{\mathrm{T}} \tag{4-25}$$

该变换可实现可逆数字水印，因为其逆变换后的元素均为整数，这也符合前面的结论。

3）基于变换矩阵的可逆数字水印实例

由上可知，将图像进行 Harr 小波变换以后，可通过修改其变换域中整数部分的直方图来实现可逆数字水印。首先将图像进行小波变换，然后利用直方图修改的方法用其高频部分（LH、HL、HH）的整数部分隐藏秘密信息，从而实现可逆数字水印。

对相邻 2×2 大小的图像块分别进行如下计算：

$$\boldsymbol{S} = \boldsymbol{H} \cdot \boldsymbol{C} \cdot \boldsymbol{H}^{\mathrm{T}} \tag{4-26}$$

式中，$\boldsymbol{H}^{\mathrm{T}}$ 表示 $\boldsymbol{H}$ 的转置矩阵。

当取参数

$$\boldsymbol{H} = \begin{bmatrix} 1/2 & 1/2 \\ 1/2 & -1/2 \end{bmatrix} \tag{4-27}$$

时，其可逆数字水印的统计结果如表 4-5 所示。

表 4-5 当变换矩阵选式(4-27)时隐藏的统计结果

载体图像	Lena	Jet	Pepper	Baboon
PSNR(dB)	44.2715	45.4938	44.1485	46.1267
嵌入率(bpp)	0.3324	0.4805	0.2752	0.1989

同理,根据该条件还可以构建 4×4 大小的变换矩阵,如式(4-28)所示。该变换矩阵同样满足变换条件,因为 $\boldsymbol{H}^{-1}$ 中的元素都为整数。

$$\boldsymbol{H}=\frac{1}{4}\begin{bmatrix}1 & 1 & 1 & 1\\ 1 & 1 & -1 & -1\\ 1 & -1 & -1 & 1\\ 1 & -1 & 1 & -1\end{bmatrix} \tag{4-28}$$

将该矩阵代入式(4-26)计算,也可实现可逆数字水印,其隐藏的结果如表 4-6 所示。

表 4-6 变换矩阵为式(4-28)时隐藏的统计结果

载体图像	Lena	Jet	Pepper	Baboon
PSNR(dB)	37.43	37.45	37.42	36.94
嵌入率(bpp)	0.5788	0.6271	0.5280	0.2597

当选取式(4-28)进行变换时,由于该变换矩阵的大小为 4×4,将其代入式(4-26),将会产生一个低频数据和 15 个高频数据。这样图像点总数的 15/16 为高频数据,直方图修改正是统计了这些数据实现了可逆数字水印。若利用式(4-27)计算可得到占总数 3/4 的高频数据,在变换后的数据中可从图像点总数的 3/4 个数据进行统计隐藏,实现可逆数字水印。由于其利用的数据量较少,所以嵌入量也相应会较少。对于较平滑的载体图像,如 Jet,其差值数据集中,所以嵌入量较大;对于不平滑的载体图像,如 Baboon,相邻像素点间的相关性相对较小,产生的高频数据就不集中,其高频的整数部分的峰值会下降,影响其嵌入率。虽然采用阶数较大的变换矩阵其嵌入率较大,但其峰值信噪比也下降较快。

4.3.7 软件水印技术

软件水印(software watermarking)是嵌入到程序当中的秘密消息,这些消息要求能够方便可靠地提取出来,以证明软件的所有权,并且具有在保证程序功能的情况下不能或者是难以去除该消息的功能。根据水印的提取技术,可将软件水印分为静态水印和动态水印(见图 4-14)。静态水印存储在可执行的程序代码中,比较典型的是把水印信息放在安装模块部分,或者是指令代码中,或者是调试信息的符号部分。对于 Java 程序,水印信息也可以隐藏在类文件(包括常量池表、方法表、行号表)的任何部分中。静态水印又可以进一步分为静态数据水印和静态代码水印。区别于静态水印,动态水印保存在程序的执行状态中,而不是程序源代码本身。这种水印可用于证明程序是否经过了迷乱变换处理。动态水印主要有 3 类:执行状态水印、数据结构水印和 Easter Egg 水印(复活节

彩蛋水印)。其中,每种情况都需要有预先输入,然后根据输入,程序会运行到某种状态,这些状态就代表水印。

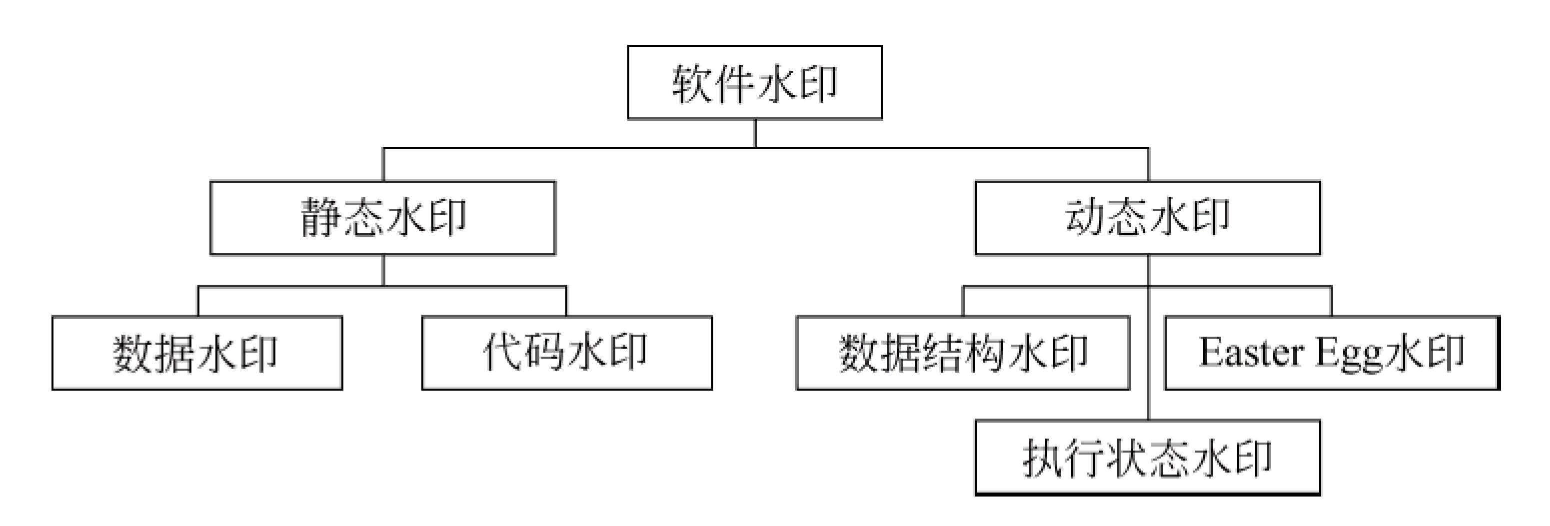


图 4-14 软件水印分类

1. 静态数据水印

静态数据水印很容易产生和识别,是一种常见的水印。这种水印可以在程序中的一些数据中体现出来,因而很容易被迷乱攻击破坏。比如把所有的数据分解成一系列数据,然后散布到整个程序中,这样代表水印信息的数据也被分解,增加了水印检测的难度;或者用一个产生这些数据的子程序来代替这些数据,这样在程序中就找不到该数据的原型,也就无法检测水印。

2. 静态代码水印

利用人类视觉和听觉的不敏感性,多媒体水印通常是加在载体上的冗余部分。也可以用相同的方式来构造代码水印,因为目标代码也包含了冗余信息。例如通过调整两条无依赖关系指令的顺序可以嵌入 1 比特的水印信息。IBM 提出了一种把寄存器出入栈的顺序作为水印的方法,同样可以通过排列有 m 个分支的 case 语句的顺序来编码 $\log(m!)$比特信息。Davidson 描述一种类似的代码水印,它在程序的控制流图的一个基本模块中对软件的序列号进行编码。

许多代码水印都经不起一些简单的水印攻击(如调整指令的顺序)。既然交换指令的顺序不影响源程序,那么就可以把源代码中所有满足这个条件的指令都交换位置,这样就无法检测到先前加入的水印了。

很多代码迷乱技术能够破坏代码水印。对于 Davidson 的方法,只要能够准确地找到控制流图的基本模块,就能很容易通过插入一个布尔值始终为 TRUE 的条件分支破坏这个基本模块,导致水印无法检测。迷乱变换会破坏所有静态结构水印。内嵌、循环变换都是常见的优化技术,但这些技术也很容易破坏静态代码水印。Moslkowitz 提出了一种具有防篡改的水印算法,其基本思想是把关键代码的一部分隐藏在软件的资源(如图标、声音)中,并且程序会不时地从资源中提取出这段代码执行,如果资源被破坏,那么程序就会出错。

静态代码水印更难抵抗语义保持变换攻击。出于安全考虑,Java 程序不能检测自己的代码,例如:

```
if (instruction # 100!= "add") exit()
```

在Java语法中是不允许的，虽然在其他语言（如C语言）中是可能的。但是由于这种语句的特殊性，它要检查指令，因此很容易找到这种语句在程序中的位置。

总之，尽管静态水印比较简单，但是由于它容易遭到破坏、鲁棒性不好，因而没有得到广泛的应用。

3. Easter Egg水印

Easter Egg水印无须检测，它通过一个输入产生一个输出。比如输入一个字符串，然后屏幕上就显示出版权信息或一幅图像。Easter Egg水印的主要问题是水印在程序中的位置很容易找到，一旦输入正确的信息，用softice这样的标准调试软件就可以跟踪程序执行情况，进而找到水印的位置，所以这种水印的安全性不高。

4. 动态数据结构水印

这种水印的原理是：输入特定信息激发程序把水印信息隐藏在堆、栈或者全局变量域等程序状态中。当所有信息都输完之后，通过检测程序变量的当前值来进行水印提取。可以安排一个提取水印信息的进程或在调试器下运行程序查看变量的取值。

与Easter Egg水印不同的是，动态数据水印没有输出，而且水印的提取过程不是封装在应用程序中，因而不容易找到水印在程序中的位置，但是这种水印也经不住迷乱变换的攻击。

5. 动态执行过程水印

动态执行过程水印是当程序在特定的输入下运行时，对程序中指令的执行顺序或内存地址的走向进行编码生成水印。水印检测则通过控制地址和操作码顺序的统计特性来进行。

软件水印是近年来才出现的软件版权保护技术，它把程序的版权信息和用户身份信息嵌入到程序中，用来标志作者、发行者、所有者、合法使用者等，并携带版权信息和身份认证信息，可以鉴别出非法复制和盗用的软件产品。它甚至被认为是数字作品内容保护的最后一道防线。

从软件水印的用途来看，有以下一些应用：

（1）软件版权申明（authorship）：通过软件水印申明软件的版权，软件中的水印信息可以被合法的用户（公开水印密钥）提取。软件用户可以通过该水印判断所使用的软件是否为正版软件。

（2）软件版权证明（authentication）：通过软件水印证明软件的版权，软件中的水印信息仅能被软件开发者（拥有水印密钥）提取，该水印信息可以证明软件的所有权。当两个公司都称软件是自己公司的软件时，软件版权证明水印可以证明软件的所有权，从而揭穿盗版者的谎言。

（3）盗版源的跟踪：在分发给不同用户的软件中嵌入的水印信息各不相同（不同的信息是软件的指纹），当盗版行为发生时，可以根据软件的指纹寻找盗版软件是从哪个用户流传出去的，从而定位盗版源。

(4) 非法复用软件模块的发现：如果整个软件被盗用，常常是很容易发现的；但当仅有某个模块被非法复用时，常常是难以发现的，软件水印可以用于发现与检测这种情况下的盗版行为。

(5) 盗版自报告：Easter Egg 软件水印利用了软件可运行的特点，把水印检测器嵌入到软件当中，当检测器运行时，可以通过检查软件的生存环境（如主机 IP 等），判断该软件的生存环境是否构成盗版行为，进而在可能的情况下通过网络主动报告盗版行为。

(6) 盗版自发现：随着计算机网络的迅速发展，通过网络分发软件成为可能。这就给软件盗版的自发现提供了可能，利用网络爬虫技术搜索网上的软件，并检测这些软件中的水印信息，从而自发地发现盗版行为。

4.4 信息隐藏与数字水印的应用与发展

4.4.1 信息隐藏技术的应用与发展方向

在信息安全领域中，信息隐藏与数字水印技术的应用可归结为下列几个方面。

1. 数字知识的产权保护

知识产权保护是信息隐藏与数字水印技术所力图解决的重要问题，信息隐藏技术的绝大部分研究成果都集中于这一领域，随着网络和数字技术的快速普及，通过网络向人们提供的数字服务也会越来越多，如数字图书馆、数字图书出版、数字电视、数字新闻等，这些服务提供的都是数字产品。数字产品具有易修改、易复制、易窃取的特点，因此，数字知识产权保护就成为迫切需要解决的实际问题，信息隐藏技术应用于版权保护时，所嵌入的签字信号通常被称作“数字水印”，数字水印技术可以成为解决此难题的一种方案。现在越来越多的视频信号、音频信号和数字图像中被贴上了不可见的标签，用以防止非法拷贝和数据跟踪服务提供商在向用户发送产品的同时，将双方的信息代码以水印的形式隐藏在作品中，这种水印从理论上讲应该是不被破坏的。当发现数字产品在非法传播时，可以通过提取出的水印代码追查非法散播者。其主要特点是版权保护所需嵌入的数据量小，对水印信号的安全性和鲁棒性要求很高。

2. 数据完整性鉴定

数据完整性鉴定是指对某一信号的真伪或完整性的判别，并需要进一步指出该信号与原始真实信号的差别，以确认资料在网上传输或存储过程中是否被篡改、破坏或丢失。假定接收到一个多媒体信号（如音频、视频或图像等），并初步判断它可能是某一原始真实信号的修改版本，数据篡改验证的任务就是在对原始信号的具体内容不可知的情况下，以最大的可能判断是否真实。

(1) 要充分利用数据库管理系统提供的数据完整性约束机制和各种输入数据的引用完整性约束设计以保证数据完整、准确地输入和储存。

(2) 在数据传输过程中可视情况选用相应的数据校验方式对传输数据进行校验

检查。

3. 数据保密

在网络上传输秘密数据要防止非法用户的截获和使用,这是网络安全的一个重要内容。随着信息技术的发展以及经济的全球化,这一点不仅涉及政治、军事领域,还将涉及商业、金融机密和个人隐私。信息隐藏技术为网上交流的信息采取了有效的保护,比如电子政务中敏感信息、电子商务中的秘密协议和合同、网上银行交易的重要数据、重要文件的数字签名以及个人隐私等,还可以对一些不愿为别人所知道的内容使用信息隐藏的方式进行隐藏储存,从而使数据得到保密,保证了信息的安全性。

4. 资料不可抵赖性的确认

在网上交易中,交易双方的任何一方不能抵赖自己曾经做出的行为,也不能否认曾经接收到对方的信息,这是交易系统中的一个重要环节。可以在交易体系的任何一方发送和接收信息时,将各自的特征标记形式使用信息隐藏技术加入到传递的信息中,这些标记应是不能被去除的,从而达到确认其行为的目的。

信息隐藏技术是近年来多媒体通信和多媒体信号处理领域中新兴的研究方向,它为信息安全提供了一种新的思路,为信息安全研究提供了一个新的方向。

目前国际上先进的信息隐藏技术已能做到隐藏的信息可以经受人的感觉检测和仪器的检测,并能抵抗一些人为的攻击。但总的来说,信息隐藏技术尚没有发展到可实用的阶段,使用密码加密仍是网络信息传输的主要安全手段。虽然目前对信息隐藏的研究有了很大的进展,在信息安全中起到了重要的作用,但仍存在大量的实际问题亟待解决,如信息隐藏的容量问题,如何建立不可感知性的数学度量模型,信息隐藏的容量上界如何计算等;信息隐藏的对立面——隐藏分析如何得到同步发展;如何对信息隐藏进行分析和分类;如何找到信息隐藏技术自己的理论依据,形成完善和科学的理论体系等。

4.4.2 数字水印技术的应用和发展方向

数字水印技术的研究大约始于1994年,已有不少著名大学和研究机构投入相当大的人力、物力和财力,致力于该项技术的研究,并取得了一定的成果,包括美国的麻省理工学院、Purdue大学、英国的Ceorage Mason大学、瑞士洛桑联邦工技院、美国的NEC研究所、美国的IBM研究所等。一些公司已推出了一些数字水印软件产品等。各研究机构正努力设计出更高效安全、更通用、更强抗攻击能力的数字水印产品。

我国在该领域的研究尚未普及,虽已引起数十家大学和研究机构(如北京邮电大学、哈尔滨工业大学、中科院自动化研究所和国防科大等)的关注,但到目前为止还基本没有成熟的技术或商业化软件可供投入市场。随着数字化产品在中国的普及,Internet在中国的迅猛发展以及电子商务的快速发展,数字水印技术将会拥有更加广阔的应用前景,这也是国产化软件走向世界的捷径。目前,数字水印还没有形成统一的国际标准。

近年来,数字水印技术的研究和发展都很迅猛,其研究方向也呈多元化,归纳起来大致有如下几个方向:

(1) 将水印处理技术与编码算法统一起来。数字产品的发布和使用通常要经过编码和传输。传统的水印处理往往与压缩编码算法分开。目前许多科研机构在研究如何在编码的过程中嵌入水印并提出了相应的嵌入方法(研究隐藏算法的同时当然也包含了检测算法)。这样做的优点在于使水印对该编码算法具有鲁棒性,尽量减少无意的水印攻击。

(2) 力图建立国际统一标准的水印处理算法。目前国际上的水印处理算法尚未形成统一的标准,形成标准已经成为所有水印研究者的共同目标。然而,由于形成国际标准的算法要求必须具有优越性、通用性、健壮性和有效性,并要得到世界各国的认同,所以形成标准是一项艰巨的任务。其中基于 DCT 变换和小波变换的水印处理技术是各国争相研究的热点,形成标准的可能性最大。

(3) 将水印处理技术商业化并应用于其他领域如军事和国防领域,用于传送秘密的军事命令、验证军事命令,信息的真实可靠性,并探索该领域的新技术和新理论,这对于国防现代化建设和未来的信息化、网络化战争有重大意义。

思 考 题

4.1 简述信息隐藏与数字水印的区别与联系。

4.2 例举常见的信息隐藏技术。

4.3 简单描述数字水印的组成框架。

4.4 针对数字水印有哪些攻击? 分别有何应对策略?

4.5 目前,可逆水印还存在哪些不足?

4.6 比较三种常见的可逆水印的方法,分析各自的优缺点。

参 考 文 献

[1] 葛陵元,胡湘陵,郑若忠. 计算机密码学. 重庆:西南交通大学出版社,1989.

[2] 宋震. 密码学. 北京:中国水利水电出版社,2002.

[3] 卢铁城. 信息加密技术. 成都:四川科学技术出版社,1989.

[4] 汪晓帆,戴跃伟,茅耀斌. 信息隐藏技术方法与应用. 北京:机械工业出版社,2001.

[5] 卢开澄. 计算机密码学——计算机网络中的数据保密与安全. 北京:清华大学出版社,2003.

[6] 王炳锡,陈琦,邓峰森. 数字水印技术. 西安:西安电子科技大学出版社,2003.

[7] 张立和,杨义先,钮心忻,等. 软件水印综述. 软件学报,2003,14(2):268-277.

[8] 刘瑞祯,谭铁牛. 数字图像水印研究综述. 通信学报,2000,21(8):39-48.

[9] 易开祥,石教英,孙鑫. 数字水印技术研究进展. 中国图像图形学报,2002,6(2):111-117.

[10] 潘蓉,高有行. 数字图像水印技术研究. 湖南大学学报(自然科学报),2002,29(2):117-123.

[11] Cox I J,Miller M L. The first 50 years of electronic watermarking. EURASIP Journal on Applied Signal Processing-Emerging applications of multimedia data hiding,2002(2),2002.

[12] Kundur D, Hatzinkaos D. Digital Watermarking for Telltale Tamper Proofing and

Authentication. Proceedings of the IEEE Special Issue on Identification and Protection of Multimedia Information,1999,87(7):1167-1180.

[13] Chang C C,Wu T C. Remote Password authentication with smart cards. proceedings of IEEE Computers and Digital Techniques,1991:165-168.

[14] Niu X M,Lu Z M,Sun S H. Digital watermarking of still images with gray level digital watermarks. IEEE Transactions on Consumer Electronies,2000,46(1):137-145.

[15] Voyatzis G,Pitas I. The use of watermarks in Protection of digital multi media Products. Proceedings of the IEEE,1999,87(7):1197-1207.

[16] Herrigel A,Voloshynovski S. Copyright and content Protection for digital images based on asymmetric cryptographic techniques. Proceedings of Sixth ACM International Multimedia Conference on Multimedia and Security Workshop,1998:99-112.

[17] Crave S. On Public-key stegnaogrpahy in the Presence of an active warden. Proceedings of Information Hiding'1998,1998,1998(1525),355-368.

[18] Hartung F,B. Girod. Fast Public-key watermarking of compressed video. Proceedings of ICIP'1997,1997,1(1):528-531.

[19] Fridrich J,Baldoza A C,Simard R J. Robust digital watermark based on key-dependent basis functions. Proceedings. of lnformation Hiding'1998,1998:143-157.

[20] Zhao J,Koch E,Luo C. Digital watermarking in business today and tomorrow. Communications of ACM,1998,41(7):67-72.

[21] Zeng W. Digital watermarking and data hiding:technologies and applications. Proceedings of ICISAS'1998,1998,3:223-229.

[22] Langelaar G C,Setyawan I,Lagendijk R L. Watermarking digital image and video data. A state-of-the-art overview. IEEE Signal processing Magazine,2000,17(5):20-26.

[23] Tian J. Reversible data emkedding using a difference expansion. IEEE Transactions on circuits and systems for video Technology. 2003,13(8):890-896.

[24] 赵彦涛.可逆信息隐藏技术及其在鲁棒数字水印中的应用研究.燕山大学博士学位论文,2010.

[25] 张立和,杨义先,钮心忻,牛少彰.软件水印综述.软件学报,2003,14(2):268-277.

[26] 刘海明.信息安全中的信息隐藏技术及其应用.科技信息,2009,4:58.

[27] 数字水印的发展状况与方向,引自 http://blog.sina.com.cn/blog-77e5b61e0/00rz7r.html.

[28] 王育民,张彤,黄继武.信息隐藏——理论与技术.北京:清华大学出版社,2006.

第 5 章 chapter 5

数字取证技术

本章学习目标

数字取证技术是当前数字内容安全的一个研究热点。本章介绍了数字取证的基本原理与相关技术，主要包括数字取证的技术分类、数字内容篡改取证、数字内容来源取证以及数字内容隐秘分析取证，并介绍了一些经典的取证案例与取证方法。

通过本章的学习，应掌握以下内容：

(1) 数字取证的基本原理。

(2) 数字取证技术的分类。

(3) 数字内容篡改取证技术。

(4) 数字内容来源取证技术。

(5) 数字内容隐秘分析取证技术。

数字取证技术是信息安全领域近年来发展起来的一个新的研究热点。它是计算机科学、法学以及刑法学等学科的交叉学科。数字取证技术的目的是调查与数字技术相关的电子商务诈骗、侵占知识产权、入侵计算机等数字犯罪，有效确保计算机、移动手机以及通信网络等数字设备中相关信息的安全，并进而构建出一个整体信息安全架构，以防止网络安全等相关攻击，协助企业、司法机构收集数字犯罪证据。

在传统的主动取证技术中，数字签名技术需要从原始数字内容中提取数字签名或内容摘要，然后通过对比接收方的数字签名与内容摘要来验证传输过程中数字内容是否经过篡改，这种方法需要事先产生辅助信息。而数字水印技术需要将数字产品的版权信息嵌入到可能存在的冗余信息中，以达到保护数字产品版权与完整性的目的，这种技术要求相关设备带有水印嵌入功能，同时要求被嵌入的水印具有较强的鲁棒性，还需要权威的三方介入，这对数字水印的应用带来了很大的局限性。而本章将要介绍的数字取证技术主要是被动取证技术，它通过对数字内容的统计特性进行分析来判断数其内容的真实性、完整性和原始性。若没有特别说明，本章所指的数字取证均为数字内容被动取证技术。

本章从数字取证的基本概念入手，首先介绍了数字取证技术的分类，接下来分别从内容篡改、内容来源及内容隐秘分析三个方面详细介绍了数字取证技术，通过本章的学习可以对数字取证技术有进一步的了解。

5.1 数字取证基本概念

5.1.1 数字取证概念

随着计算机及网络技术的高速发展和广泛应用，利用计算机进行犯罪也在日趋增加。要想遏制这类犯罪案件的发生，就需要能证明犯罪的证据，从计算机中提取证据成为案件侦破的关键。计算机取证对于起诉这类犯罪行为至关重要。

计算机犯罪取证(数字取证)也被称为计算机法医学，是指把计算机看做犯罪现场，运用先进的辨析技术，对电脑犯罪行为进行法医式的解剖，搜寻确认罪犯及其犯罪证据，并据此提起诉讼。它作为计算机领域和法学领域的一门交叉科学，正逐渐成为人们关注的焦点。

数字取证是指为了揭示与数字产品相关的犯罪或过失行为，以及由其他原因导致的使系统发生故障的现象，利用一切科学合法的方法和工具，对以 0/1 二进制表示的数据进行识别、保存、收集、检查、分析和呈堂等活动过程。数字取证是个广义的范畴，从其研究范围来讲，既包括计算机取证又包括网络取证；从其内涵来讲，是对数字资源的提取、存储、分析和利用，它与网络取证和计算机取证的本质是一致的。

数字取证的对象是电子证据。电子证据不同于其他证据形式，是指以电子的、数字的、电磁的、光学的或类似性能的相关技术形式保存记录于计算机、磁性物、光学设备或类似设备及介质中或通过以上设备生成、发送、接受的能够证明刑事案件情况的一切数据或信息，属于高科技证据。我国《刑事诉讼法》规定证据有 7 种形式，即物证、书证；证人证言；被害人陈述；犯罪嫌疑人、被告人供述和辩解；鉴定结论；勘验、检查笔录；视听资料。在这 7 种证据中没有电子证据这种证据形式。因此，电子证据需经过法律规定的收集和审查才能具备证据能力和证据性。如何以可见、可感知和可移动的形式将电子证据固定下来，从而在技术上实现电子证据的有形性和可视性的转变是数字取证的重要过程。这种新形式的证据与传统刑事证据相比有许多不同的特点，具体如下：

(1) 数字化特性。计算机内的文档、图形、图像、动画、音频、视频等信息形式均是以二进制数据格式存储、传输。

(2) 电子介质特性。电子证据生成后存储于计算机硬盘、软盘、光盘、磁带等电子设备及介质中。

(3) 具有较强的隐蔽性。计算机证据在计算机系统中存在的范围很广，使得证据容易被隐藏。一切信息都由编码来表示并传递，使得计算机证据与特定主体之间的关系按照常规手段难以确定。

(4) 客观实在易变性。计算机数字信息的存储和传输过程中又容易被截取、监听、剪接、删除，同时还可能由于计算机系统、网络系统、物理系统的原因，造成其变化且不留痕迹。

(5) 取证的广域性。计算机犯罪实施可以在计算机网络中延伸到世界范围的任何一个角落发生。网络的便利性使得计算机网络犯罪跨越省界、国界都是很容易做到的，这

给数字取证工作带来很大的挑战。

数字证据与传统的证据相比较，有以下突出的特点：

(1) 数字证据同时具有较高的精密性和脆弱性。一方面，数字证据以技术为依托，很少受主观因素的影响，能够避免其他证据的一些弊端，如证言的误传、书证的误记等；另一方面，由于数字信息是用二进制数据表示的，以数字信号的方式存在，而数字信号是非连续性的，故意或因为其他差错对数字证据进行的变更、删除、删节、剪接、截获和监听等，从技术上讲很难查清。

(2) 数字证据具有较强的隐蔽性。数字证据在计算机等数字系统中可存在的范围很广，使得证据容易被隐藏。另外，由于数字证据在存储、处理的过程中，其信息的表示形式为二进制编码，无法直接阅读。一切信息都由编码来表示并传递，使得数字证据与特定主体之间的关系按照常规手段难以确定。

(3) 数字证据具有多媒体性。数字证据的表现形式是多样的，尤其是多媒体技术的出现，更使数字证据综合了文本、图形、图像、动画、音频及视频等多种媒体信息，这种以多媒体形式存在的数字证据几乎涵盖了所有的传统证据类型。

(4) 数字证据还具有收集迅速、易于保存、占用空间少、容量大、传送和运输方便、可以反复重现、便于操作等特点。数字证据的这些特点表明数字取证面临不少难题，有完全不同于传统取证的问题需要研究。数字取证与国家安全、司法安全以及国防安全密切相关，并已成为信息安全领域的研究热点之一。由于其本质和信息安全学科之间存在差别，国外学者已提出建立数字取证新学科，并研究了教育与研究领域的人才培养体系与知识结构。据了解，国内有些学者也在关注数字取证作为学科发展的新领域。

5.1.2 取证过程模型

美国国家司法研究所(U.S. National Institute of Justice，NIJ)2001年公布了关于数字犯罪现场调查的过程模型，其目的在于提供有关数字犯罪现场调查指导方针，以用于指导调查。该模型主要针对于调查人员在首次调查数字犯罪过程中当遇到不同类型的数字证据时，给予相应的处理程序，从而可以更加安全地处理相关的数字证据，其重点在于数字调查中的收集过程。该模型如图5-1所示。

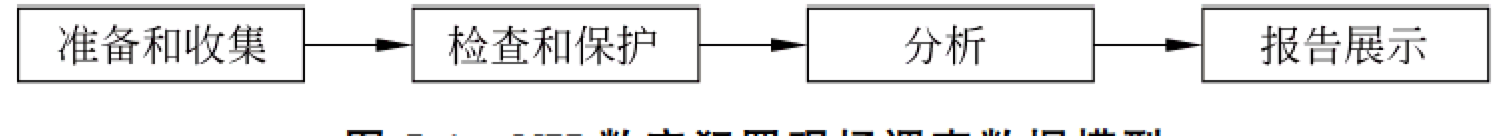


图 5-1 NIJ 数字犯罪现场调查数据模型

该模型包括如下要点。

1. 电子证据的确定和收集

要保存计算机系统的状态，避免无意识破坏现场，同时不给犯罪者破坏证据提供机会，以供日后分析。包括封存目标计算机系统并避免发生任何的数据破坏或病毒感染，绘制计算机犯罪现场图、网络拓扑图等，在移动或拆卸任何设备之前都要拍照存档，为今后模拟和还原犯罪现场提供直接依据。在这一阶段使用的工具软件由现场自动绘图软

件、检测和自动绘制网络拓扑图软件等组成。

获取证据从本质上说就是从众多的未知和不确定性中找到确定性的东西。这一步使用的工具一般是具有磁盘镜像、数据恢复、解密、网络数据捕获等功能的取证工具。

2. 电子证据的保护

这一阶段将使用原始数据的精确副本，应保证能显示存在于镜像中的所有数据，而且证据必须是安全的，有非常严格的访问控制。为此必须注意以下几点：

(1) 通过计算副本和原始证据的Hash值来保证取证的完整性。

(2) 通过写保护和病毒审查文档来保证数据没有被添加、删除或修改。

(3) 使用的硬件和软件工具都必须满足工业上的质量和可靠性标准。

(4) 取证过程必须可以复验。

(5) 数据写入的介质在分析过程中应当写保护，以防止被破坏。

(6) 分析检查阶段的证据，以确定“重要性和证据力”。

(7) 在每个案件之后，创建检查日志记录。

3. 电子证据的分析

具体包括文件属性分析技术、文件数字摘要分析技术、日志分析技术、密码破译技术等。分析阶段首先要确定证据的类型，主要可分为三种：

(1) 使人负罪的证据，支持已知的推测。

(2) 辨明无罪的证据，同已知的推测相矛盾。

(3) 篡改证据，以证明计算机系统已被篡改而无法用来作证。

4. 报告展示阶段

给出调查所得结论及相应的证据，供法庭作为公诉证据。还要解释是如何处理和分析证据的，以便说明监管链和方法的彻底性。

该模型主要目标是收集阶段，因为检查和分析阶段仅仅给出了可能含有某一类型犯罪的证据的数据类型，而没有详细列出其他的细节。此外，检查和分析阶段对应的需求区别并不明显。因为在检查阶段，使用数据约简技术只是后续分析阶段中一种比较普通的技术，换言之，在分析阶段，可以执行数据约简技术来识别重要的证据。因此在模型中同时包含这两个处理过程是有争议的。之外，如果是这种情况的话，那么这两个阶段可以合并成一个含有分析技术的阶段。

随着数字取证技术的发展，人们逐渐关注数字取证中更为本质的内容，出现了抽象的取证模型，如数字取证研究工作组(Digital Forensics Research Workshop，DFRWS)的取证框架以及提出的抽象过程模型等。之后，为了进一步完善取证模型，产生了将物理犯罪调查与数字取证调查进行结合的抽象模型、针对于安全事件的取证调查模型、针对于调查目标的取证模型以及端到端取证模型等。这些工作有力地推动了数字取证技术的发展，对取证标准化具有比较大的意义，为相关的立法工作也提供了支持。

5.1.3 数字取证常用工具

计算机取证技术也日益成熟，各种计算机取证软件、计算机取证工具层出不穷，仅仅针对逻辑层的就有 Guidance 的 Encase、AccessData 的 FTK、FINALData 的 FINALForensics 等诸多软件，针对物理层的计算机取证工具也不胜枚举，但是要达到更有效打击计算机犯罪的目的，法证界迫切需要多元化的计算机取证综合解决方案。

计算机取证的相关工具包括一般工具软件，如用于检测分区的工具软件、杀毒软件、各种压缩工具软件等。还有取证专用工具软件，如文件浏览器、图片检查工具、反删除工具、CD-ROM 工具、磁盘擦除工具等。

Encase 自称是唯一一个完全集成的基于 Windows 界面的取证应用程序，是专业的计算机取证工具，包括 Encase 取证版解决方案和 Encase 企业版解决方案。

Encase 取证版解决方案是国际领先的受法院认可的计算机调查取证的工具。具有以下主要特性：

(1) 支持并能管理易变的时区。

(2) 能分析 UNIX 和 Linux 的系统文件。

(3) 能查看并搜索 NTFS 压缩文件，能检测 NTFS 文件系统中的附加分区中的信息。

(4) 允许查看 NTFS 文件/文件夹的所有者和访问权。

(5) 允许用户限制其可查看的数据，并能保护特权数据。

(6) 具有良好的 EnScript 程序界面，编辑和调试代码操作更方便。

(7) 可以隐藏用户定义的扇区或提前读取一定数量的扇区，从而提高导航函数的速度。

(8) 具有多个关键词搜索算法，能够动态加快搜索速度。

(9) 支持 RAID，了解动态磁盘分区结构并能处理所有可能的配置。

Encase 企业版解决方案由 SAFE、Examiner 和 Servlet 三部分组成，是世界上第一个可有效执行远程企业紧急事件响应(response)、审计(audit)和发现(discovery)任务的解决方案。

5.2 数字取证分类

5.2.1 数字取证技术的分类

从计算机取证技术的发展来看，先后有数字取证(digital forensics)、电子取证(electric forensics)、计算机取证(computer forensic)、网络取证(networks forensics)等术语。

1. 电子取证

随着计算机犯罪个案数字不断上升和犯罪手段的数字化，搜集电子证据的工作成为

提供重要线索及破案的关键。恢复已被破坏的计算机数据及提供相关的电子资料证据就是电子取证。

电子取证主要研究除计算机和网络以外的电子产品中的数字证据获取、分析和展示,如数码相机、复印机、传真机甚至有记忆存储功能的家电产品等。

2. 计算机取证

计算机取证的主要方法有对文件的复制、被删除文件的恢复、缓冲区内容获取、系统日志分析等,是一种被动式的事后措施,不特定于网络环境。

3. 网络取证

网络流的相关性、数据的完整性和包捕获的速率是网络取证分析首要考虑的事情。相关性是指在某些环境下,应当在捕获网络流时应用过滤器去掉不相关的数据。数据的完整性要求网络取证工具应当一直监控网络流。

网络取证对数据的保护和一般的数字取证过程要求相同,网络取证分析的相关技术包括人工智能、机器学习、数据挖掘、IDS 技术、蜜阱技术、SVM 和专家系统等。

根据网络攻击一般过程,网络取证模型如图 5-2 所示。

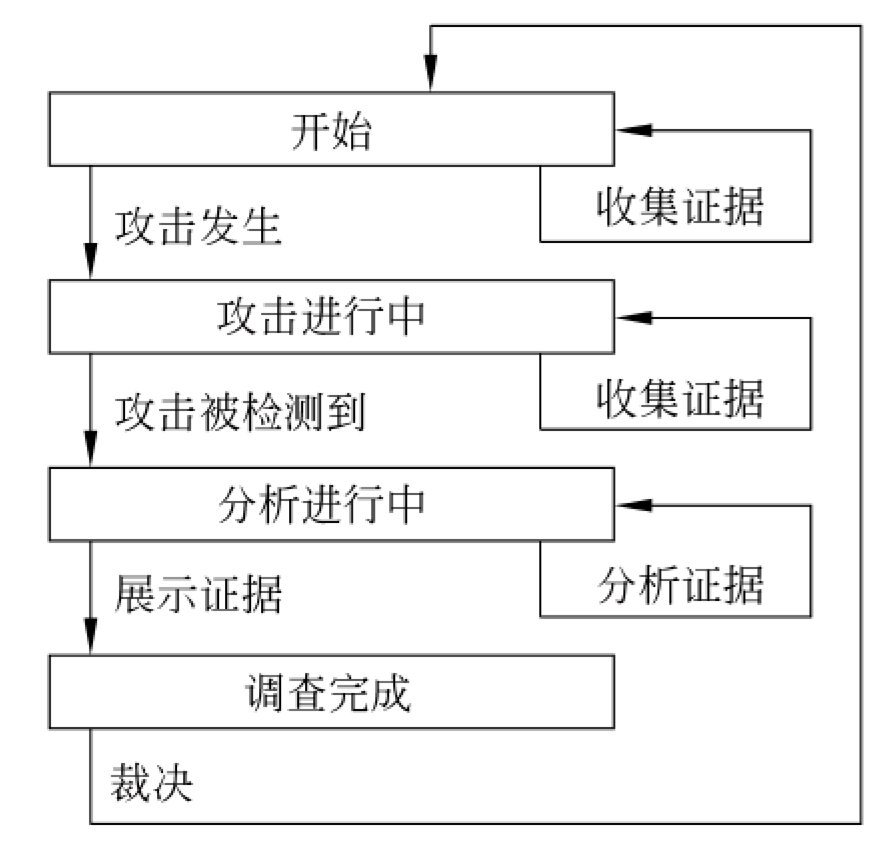


图 5-2 网络取证模型

下面简单介绍几种常见的网络取证技术。

1) IDS 取证技术

将计算机取证结合到入侵检测等网络安全工具和网络体系结构中进行动态取证,可使整个取证过程更加系统并具有智能性和实时性,并且还能迅速做出响应。IDS 取证的具体步骤如下:

① 寻找嗅探器(如 sniffer)。

② 寻找远程控制程序。

③ 寻找黑客可利用的文件共享或通信程序。

④ 寻找特权程序。

⑤ 寻找文件系统的变动。

⑥ 寻找未授权的服务。

⑦ 寻找口令文件的变动和新用户。

⑧ 核对系统和网络配置,特别注意过滤规则。

⑨ 寻找异常文件,这将依赖于系统磁盘容量的大小。

⑩ 查看所有主机,特别是服务器。

⑪ 观察攻击者,捕获攻击者,找出证据。

⑫ 如果捕获成功则准备起诉,如立刻联系律师等。

⑬ 做完全的系统备份,将系统备份转移到单用户模式下,在单用户模式下制作和验

证备份。

2）蜜阱取证技术

蜜阱是包括蜜罐和蜜网等以诱骗技术为核心的网络安全技术。它是一种经过精心设计的诱骗系统，当黑客进行攻击时，它能够监视攻击者的行径、策略、工具和目标，从而自动地收集相关的电子证据，实现实时的网络取证。

利用蜜阱进行取证分析时，一般应遵循如下原则和步骤：

(1) 确定攻击的方法、日期和时间(假设 IDS 的时钟和 NTP(网络时间协议)参考时间源同步)。

(2) 尽可能多地确定有关入侵者的信息。

(3) 列出所有入侵者添加或修改的文件，并对这些程序(包括未编译或未重组部分，因为这些部分可能对确定函数在此事件中的作用和角色有帮助)进行分析。

(4) 建立一条事件时间线，对系统行为进行详细分析，注意确认证据的来源。

(5) 给出适合管理层面或新闻媒体需要的报告。

(6) 对事故进行费用估计。

3）模糊专家系统取证技术

Jun-Sun Kim 等人开发了一个基于模糊专家系统的网络取证系统，它由六个组件组成。

(1) 网络流分析器。完成网络流的捕获和分析，为了保证数据的完整性，它要求捕获所有的网络流。分析器应用规则对捕获到的网络流进行重组，这种分类数据包的规则是协议相同的和时间连续的。

(2) 知识库。存储模糊推理引擎所使用的模糊规则，其形式为：

如果

$$X_1 = A_1, X_2 = A_2, \cdots, X_n = A_n$$

则

$$Y = Z$$

(3) 模糊化。确定每个语义变量的模糊集所定义的隶属函数和每个模糊集中输入值的隶属度。

(4) 模糊推理引擎。当所有的输入值被模糊化为各自的语义变量后，模糊推理引擎访问模糊规则库，进行模糊运算，导出各语义变量的值。

(5) 反模糊化。运用“最小-最大”运算产生输出值，作为取证分析器的输入。

(6) 取证分析器。判断捕获的数据包是否存在攻击，它的主要功能是收集数据、分析相关信息，并且生成数字证据。

4）SVM 取证技术

SVM(Support Vector Machine)取证技术是为了发现信息行为的关键特征，去除无意义的噪声，有助于减少信息存储量，提高计算速度等。同时，网络取证应该是主动的防御，对未知的网络攻击具有识别和取证能力。SVM 特征选择的基本思想是：

(1) 选择训练集和测试集，对每个特征重复以下步骤。

(2) 从训练集和测试集中删除该特征。

(3) 使用结果数据集训练分类器。

(4) 根据既定的性能准则,使用测试集分析分类器的性能。

(5) 根据规则标记该特征的重要性等级。

5) 恶意代码技术

恶意代码指能够长期潜伏、秘密窃取敏感信息的有害代码程序,应用同样的原理,可以设计用来进行取证。

5.2.2 证据取证分析技术分类

数字取证技术是指在取证调查过程中,在相关理论指导下,使用合法、合理、规范的技术或手段,保证针对计算机等数字设备取证的正确进行,同时产生真实、有效的结论。然而,目前的取证调查技术多数是为解决数字取证调查中的实际问题而发展起来的技术,没有进行充分的验证,缺乏相应的理论基础,从而在确定技术标准方面存在差异。取证分析技术可以分为以下三类。

1. 基于取证过程模型的分析技术

基于取证过程模型的分析技术的基本思想是根据 DFRWS 给出的取证过程模型进行划分的。按照该方法,取证分析技术可以分为六类:识别类、保存类、收集类 、检查类、分析类以及出示类。这种技术分类的初衷是从数字取证调查过程的角度来进行分类。这种分类的不足在于:由于缺乏相关理论指导,所以不能涵盖所有的取证分析技术种类,比如文件系统取证分析技术是数字调查的重要分析技术,它包含 NTFS 文件系统取证分析、FAT 系列文件系统取证分析以及移动设备文件系统(Symbian、Android 等文件系统)取证分析等,但是在这个分类体系中却没有说明;还有有害代码取证检测技术等。

2. 基于数字设备运行历史模型的取证分析技术

基于数字设备运行历史模型的取证分析技术最早由 Brain Carrier 提出,他从数字设备的运行历史角度来对数字取证分析技术进行分类。其主要思想是:计算机等数字设备在运行中包含一个历史过程,该过程中存在事件和状态的序列。因此在数字取证过程,将根据事件和状态的序列集合进行分析。按照数字设备运行历史模型,可以将数字取证分析技术分为七大类,并将其进一步分为 31 类分析技术。这七大类分析技术是:通用调查过程、历史周期、原子存储系统配置、原子事件系统配置、原子状态和事件定义、复杂存储系统配置以及复杂事件系统配置。

3. 基于存储介质的取证分析方法

这里的存储介质主要包括硬盘、光盘、软盘、U 盘、内存以及其他形式的存储介质。该方法围绕存储介质中证据的获取、保护、传输以及分析等进行取证调查。按照介质中数据的生命周期,该方法可以分为两类:基于永久性存储介质的取证分析和基于易失性内存的取证分析方法,前者的典型代表是磁盘取证,后者是内存取证等,其具体分析过程依赖于存储介质中的文件系统结构、原理以及内存中的进程结构等。

以上证据取证分析技术分类之间有联系、有交叉，同时也存在互补关系，且任何一类技术都具有相对性。此外，从取证分析技术发展及其应用角度看，新的分析技术也在不断出现，如文件雕刻取证技术研究已经成为下一代取证研究的重点，且目前已经获得了初步研究成果，如文档碎片分类、重组等研究。该技术同现有的取证技术结合可以更加有效解决数字取证调查问题。

5.2.3 取证技术产品、标准和规范

法律实施部门迫切需要保证数字取证工具的可靠性，即要求取证工具能稳定地产生准确和客观的测试结果。然而，目前的数字取证领域中有大约150个取证工具，其中很少是根据取证标准和规范进行研制的，甚至连比较有名的专业取证软件产品，如NTI(New Technology Inc)开发的取证产品ENCASE等是根据取证实践经验进行研制的。许多开源的取证软件产品，如dd、TCT、The Sleuth Kit等产品，也是如此。

为了获取更加具有法律效力的取证结果，美国国家标准与技术研究院(National Institute of Standards and Technology，NIST)指定了计算机取证工具测试计划(Computer Forensic Tool Testing，CFTT)，其目标是通过开发通用的工具规范、测试过程、测试标准、测试硬件和测试软件，以建立用于测试计算机取证软件的方法。该测试方法是基于一致性测试和质量测试的国际方法，符合ISO/IEC 17025:1999(能力测试和校准实验室)的一般要求。

目前，该计划组已经完成了硬盘写保护软件测试标准的制定，正在制定磁盘映像软件的测试标准，进一步将制定被删除文件恢复软件的测试标准。显然，CFTT为数字取证标准化的探讨和实践提供了一个良好的开端，有效地促进了取证产品的行业标准和规范的制定工作。

5.3 数字内容篡改取证

5.3.1 数字内容篡改手段

针对数字内容篡改，我们将分别从数字图像、数字音频和数字视频等方面对其篡改手段进行介绍。

1. 数字图像的篡改手段

(1) 图像合成：图像合成(composition)是将对象从图像背景中分离出来，添加到另一个图像背景中重新组合，以构成一幅新的画面，它包括复制-粘贴(copy-paste)与图像拼接。图像合成可以用来隐藏原始图像中的重要目标。图像合成篡改是数字图像真实性篡改最常见的方法。在现实应用中，由于图像合成篡改中所应用的两幅和多幅图往往在分辨率、合成物体的比例大小、位置等的不同，图像合成往往需要和一些其他的图像处理手段，如图像缩放、旋转、润饰等相结合起来以达到更好的篡改效果。数字图像复制-粘贴

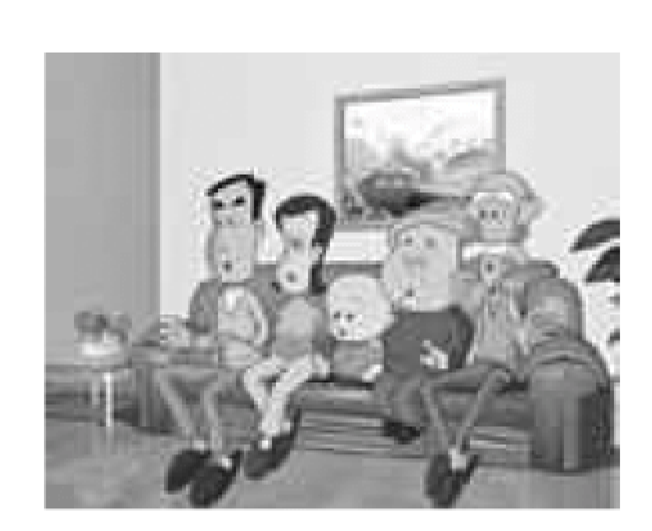
篡改是指把图像中某区域的图像内容复制粘贴到同幅图像或异幅图像的另一区域，达到与目标对象同时在场或隐藏某目标对象的目的，如图 5-3 所示。图像拼接是指对一幅或多幅图像进行裁剪，将目标对象拼接到一块，然后对其进行模糊、缩放、旋转等后处理，使得篡改痕迹不易被人察觉，从而形成原本相互独立的两个或多个场景同时在场的拼接图像，如图 5-4 所示。



(a) 背景空间



(b) 动画人物



(c) 背景空间与动画人物的合成



图 5-3　图像间的复制-粘贴



(a) 场景Ⅰ



(b) 场景Ⅱ



(c) 场景Ⅲ



(d) 场景Ⅰ、场景Ⅱ和场景Ⅲ的合成图像

图 5-4　图像拼接

(2) 图像增强：图像增强(enhancement)是指为了改善图像的视觉效果，对图像特定区域的颜色、灰度、亮度或对比度等属性进行的适当调整，这种操作能够增强图像的整体或局部特性，使原本不清晰的图像变得清晰，或突出感兴趣区域，抑制不感兴趣区域，从而达到加强图像识别效果的目的。

(3) 图像润饰：图像润饰(retouch)是指针对图像的一种修补操作，它的目的通常是为了消除某种痕迹，达到美化图像的效果。如在照相馆中，摄影师会通过润饰操作消除脸上的皱纹和黑斑。此外，润饰还经常用于图像篡改后的处理，如对复制-粘贴后的图像采用模糊润饰来消除拼接的痕迹。

(4) 图像修复：图像修复是指从图像原有信息的角度出发，对图像中的信息丢失区

域进行修复,使观察者无法识别出信息丢失区域的真伪。基于样本纹理合成的修复技术是图像修复的典型方法,该技术可以用来无痕消除图像中的大目标。

(5) 图像变形:图像变形(morphing)是一种把一幅图像逐渐变为另一幅图像的技术。采用的方法一般是分别找出原图像和目标图像上对应的特征点,然后以不同的权重叠加两幅图像,这样得到的图像就兼有两幅图像的特征,此类操作常用于动画设计和计算机生成图像的制作中,如图 5-5 所示。



图 5-5　图像变形

(6) 计算机生成:计算机生成图像(Computer Graphics,CG)是由计算机软件(如 3Ds Max、Maya 等)生成,它包括 Photorealistic CG(PRCG)和 No-Photorealistic CG(NPRCG)两类。对于 PRCG 而言,从视觉上来看与真实场景图像已很难从肉眼上区别开来。计算机生成图像是指利用计算机和图形处理软件生成现实中根本不存在的场景图像,如图 5-6 所示。计算机生成图形时,首先需要根据期望值模拟出一个三维的多边模型,然后对这个模型进行纹理和颜色修饰,修饰完成后将该模型送到模拟光源下的虚拟照相机前成像即可。计算机生成的图像在平滑度、直方图、色彩和纹理复杂度等方面与自然场景拍摄的图像会有很大的区别。

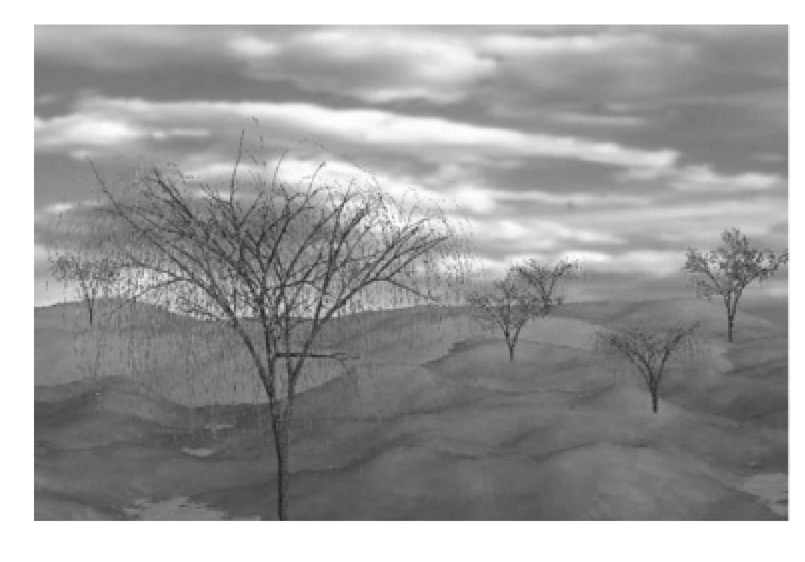




图 5-6　计算机生成图像

(7) JPEG 重压缩:JPEG 重压缩是指将 JPEG 格式的原始图像解压到空域中进行图像合成篡改后再次将图像保存为 JPEG 格式的过程,在这个过程中图像会经历一个不可逆的双重 JPEG 有损压缩过程,同时也会在 JPEG 重压缩图像中引入单次 JPEG 压缩所没有的特征。

(8) 二次获取:二次获取图像是指利用数码相机或摄影机等设备将自然场景图像拍摄生成一次图像后,又由照片的照片、照片扫描等形式生成的二次图像。

图像篡改技术各式各样,但是在上述几种常见的图像篡改手段中,伪造者用得更多的主要是复制-粘贴(合成)操作、模糊润饰操作和 JPEG 重压缩操作。

2. 数字音频的篡改手段

(1) 数字音频片断的删除：音频文件在编辑软件中的编辑窗口显示的是连续的波形，从波形上就可以确定相关语音片段的起始和终止的位置，在保证语义的连续性一致下，可以将其中一些语音片断删除，剩余部分会自动连接在一起，导致整个数字音频的语义发生变化或者掩盖关键语义。

(2) 数字音频片段的剪接：从保证语义的连续性出发，可以很简单地从各个素材中把语义片断剪辑出来，然后再拼接在一起，形成一段新的语音，达到创造新语义的效果。

(3) 数字音频片段的插入：从保证语义的连续性出发，可以在音频片段的空隙处插入剪辑片段以曲解破坏原有语义形成一段新的语音。

(4) 数字音频片段的叠加：在音频编辑软件中可以把几条声轨的内容混合在一起，意味着在某些场合能达到掩盖真声的效果或者为并不同时发生的场景添加参照物以伪造现场。

(5) 修饰局部的音频片段：针对关键的语音片段，可以做到添加噪声，使得原先的声音听上去变得模糊；或者做一些变频处理，使得说话者的声音听上去像另外一个人的声音。这些篡改的技术手段就比较高级，需要有对软件使用的经验或者要求需要有专业人士才拥有的技术手段。

3. 数字视频的篡改手段

一个标准的视频篡改过程一般都要经过以下三个操作。

(1) 帧复制或插入、删除操作，即视频合成操作。

(2) 模糊润饰操作，即为了消除局部复制造成的可视篡改痕迹的操作。

(3) MPEG 二次压缩操作，在完成了像素域的篡改操作后，需要对图像序列进行重新压缩。

此外数字图像和数字音频中的篡改手段也可以应用于数字视频。

5.3.2 数字内容篡改取证方法的评价指标

篡改检测主要是为了解决数字多媒体数据的完整性和原始性鉴定问题。传统的数字水印技术可以作为篡改检测的一种手段，但在现实世界中，绝大多数多媒体数据没有嵌入水印信息，因此，依赖水印的方法不太现实。另一方面，任何形式的篡改操作都会不可避免地引起多媒体数据内部特征尤其是统计特征的变化，由此可以借助不需要外部嵌入信息的数字取证方法来实现篡改检测。不管是何种数字内容篡改取证，其评价指标都是看其是否证明了数字多媒体数据是否完整、数据是否是原始的等问题。

数字内容篡改取证方法的几种较为常见的评价指标为：准确度(precision)、完整度(recall)和 F_1-measure。假设 P 和 R 分别表征了检测算法的准确度和完整度。其定义公式分别如下：

$$P = \frac{\text{实警率}}{\text{实警率} + \text{虚警率}} \tag{5-1}$$

$$R=\frac{\text{实警率}}{\text{实警率}+\text{错误否定率}} \tag{5-2}$$

从式(5-1)和式(5-2)可以看出，高的检测正确率需要有低的虚警率(false positive)，而高的完整度需要有低的错误否定率(false negative)。其中，检测正确率主要由正确检测区域的所占的概率所决定，因此检测正确率的定义可以直接由正确检测的区域来度量。而检测完整度主要是由于正确的操作区域被检测出来的概率来度量的，因此检测完整度的定义可以直接由检测区域中检测出的已变化区域占总的变化区域的比重来度量。则式(5-1)和式(5-2)可以变为：

$$P=\frac{\text{篡改区域}\cap\text{检出区域}}{\text{检出区域}}\times 100\% \tag{5-3}$$

$$R=\frac{\text{篡改区域}\cap\text{检出区域}}{\text{篡改区域}}\times 100\% \tag{5-4}$$

然而以检测正确率和检测完整度来衡量检测算法仍然较为片面，且提高检测正确率会降低检测完整度，反之亦然，因此两者的评价效果不能达到很好的直观性。将两个指标合并，形成 F_1-measure：

$$F_1\text{-measure}=\frac{2}{\frac{1}{P}+\frac{1}{R}}=2\frac{PR}{P+R} \tag{5-5}$$

以数字图像为例，在数字图像篡改检测中，一般将不同图像之间的篡改检测算法归纳到图像篡改检测的盲检测中，通过对图像特征的提取及分类建立算法，识别出篡改图像和自然图像。此类算法通常采用两种方法对算法分类性能做评价：一种是常用的评价参数 TP、TN、Accuracy。其中 TP、TN 分别表示真实、篡改图像的分类正确率，Accuracy 则是算法中最为关注的总体检测正确率。而另一种是用 AUC(Area Under the ROC Curve)表示 ROC(Receiver Operation Characteristics Curve)曲线下的面积来衡量分类效果。AUC 越大，则分类效果越好。

5.3.3 数字内容篡改取证方法

1. 数字图像篡改取证方法

数字图像篡改取证方法大致可归纳为以下五类。

1) 基于像素的检测方法

通过检测像素级别上的统计异常信息可判断图像是否经过篡改。针对最为常见的复制-粘贴篡改类型，可以采用搜索图像中有无完全相同的区域，其中有些是通过比较离散余弦变换块的系数，有些是通过比较图像块的主元分量。这类方法的原理很简单，关键是如何提高块的搜索效率以及抵抗由加性噪声和有损压缩而引起的图像像素的轻微变化。还有提出依据重采样所导致的特殊周期性进行篡改检测。复制-粘贴篡改往往伴随有缩放、旋转和拉伸等操作，而缩放、旋转和拉伸操作可以看成是图像信号向上和向下采样的组合，即发生多重采样，这会在图像信号中留下重采样痕迹，使图像中像素与其周围像素之间产生周期性的相关性。针对拼接合成篡改操作，可以利用像素的高阶统计特

性进行检测，特征向量由图像质量的评价测度和统计矩特征量联合构成。比较常见的方法主要使用以下三类取证特征，即图像质量特征、二值相似度测度以及高阶小波系数统计特征，其中图像质量特征又包括基于像素差的测度、基于相关性的测度、基于边缘的测度、基于人类视觉特征的测度和基于频谱距离的测度，然后分为透视、半盲和全盲三种模式来讨论篡改问题。在透视模式下，篡改的类型和强度都已知，可比较各种方法对于篡改操作的敏感度；在半盲模式下，已知篡改的类型，可通过比较上述三类测度的改变来确定篡改强度；在全盲模式下，篡改的类型未知，通过设计不同的盲分类器来实现对不同类型篡改的检测。

2）基于压缩格式的检测方法

取证的首要准则是保护证据，从这层意义上说，有损图像压缩方案可能是取证分析的最大障碍。然而有损压缩所具有的独特特性还可被用于取证分析。JPEG 是使用最普遍的图像压缩格式，检查 JPEG 图像篡改有两个主要的途径：双重 JPEG 压缩和 JPEG 的块效应。通常，原始图像和篡改后的图像都用 JPEG 格式保存，尽管双重 JPEG 压缩不一定表示图像被篡改，但这类图像有被篡改的嫌疑。通过分析离散余弦变换（Discrete Cosine Transform，DCT）系数的直方图在单次和两次压缩下的不同，可以通过两种方法估计第一次压缩时所使用的量化系数。第一种方法是利用不同量化因子进行穷举试探；第二种方法利用神经网络分类器进行分类。前一种方法计算量大，而后一种方法计算量相对较小。在一定条件下，双重压缩后 DCT 系数的直方图上会存在周期性的噪声，利用 DCT 系数直方图的傅里叶变换可以估计出第一次压缩所使用的质量因子。JPEG 二次压缩的检测是通过将 JPEG 图像中当前像素与其四邻的差值构成一个新的二维矩阵，并用一步 Markov 随机过程来描述这个差值矩阵。由于二次 JPEG 压缩减弱了上述差值矩阵中元素之间的相关性，所以可通过分析差值矩阵中元素相关值的分布来确定是否发生二次 JPEG 压缩。除了利用双重 JPEG 压缩的特征外，JPEG 的块效应是否遭到破坏也被广泛用于篡改检测。可以通过引入一个块效应特征矩阵来反映未经剪切或再压缩图像的对称性，并指出这个对称性在遭到剪切或再压缩后会被破坏。此外，还可利用 DCT 系数直方图的能谱在图像修改前后的二阶差分的极小值来估计量化系数，然后通过计算并比较各块噪声测度确定是否发生篡改以及发生篡改的位置。

3）基于成像设备特性的检测方法

由于数码相机的镜头、成像传感器和数字信号后处理会在成像过程中留下特有的设备痕迹和噪声，可以通过检查设备痕迹和噪声的一致性来判断是否发生篡改。例如，一幅自然图像内的色彩偏差应该是一致的，而篡改操作会破坏这种一致性。所以可以根据色彩偏差的一致性判断图像是否发生篡改。又如，由于目前大部分数码相机只有一片 CCD 或 CMOS 成像传感器，所获得的彩色图像都是借助颜色滤波器阵列（Color FilterArray，CFA）的插值运算得到，而不同数码相机采用的插值方法存在差异。常见的插值种类包括双线性插值、双三次插值、基于色调缓慢变换的插值、根据梯度判断边缘走向的插值以及基于自适应原则的插值等，所有这些插值运算都会在图像的各个色彩通道内、像素间引入特殊的周期性的统计相关性。可以通过检测插值像素的周期相关性是否

被破坏判断图像是否经过篡改。也可以采用检测周期性的最大似然估计迭代算法——EM(Expectation/Maximization)算法进行。还可以通过检查图像内相机响应函数(Camera Response Function,CRF)的一致性来判别图像的篡改历史。此外,还可将相机的成像环节中噪声模型参数的一致性作为图像篡改的检测依据,例如从图像去噪、小波系数分析和邻域预测等三个方面提取统计特征,利用支持向量机对统计噪声特征分类来确定图像是否发生篡改。

4) 基于物理原理的检测方法

光照条件尤其适合于检测拼接-合成类型的篡改图像。通过检测物理对象、光线和相机在三维空间中两两交互作用之间的异常可以判断图像是否发生篡改。图像篡改可归结为对图像内容的增加、删除、更改操作,一般是将一幅图像中的对象或背景与另一幅图像的背景或对象重新组合形成伪造图像,或是删除图像中的某一对象或背景来隐藏重要的目标。这些操作通常会破坏自然图像的光照一致性,而篡改操作很难把光照效果和定向的光源相匹配,因此,可根据图像中场景的光照不一致性鉴别图像的篡改。基于光学原理检测方法的关键是建立物理对象、光线和相机之间的光照模型。已有研究在单光源下二维和三维光照模型下对多光源复杂环境下的成像进行了讨论,并给出了一个复杂光源环境的低参数近似模型。

5) 基于对象几何关系的检测方法

照相机中心在图像平面上的投影点称为"主点"。在所拍摄的图像中,主点位于图像中心附近。当图像中的人或物平移时,主点也成比例地平移。通过检验从图像的不同局部所估计出的主点位置是否一致来判断图像内容是否经过了改动。

总之,图像篡改检测方法中前三类方法的理论基础是数字图像处理、信号处理和模式识别,而后两类的理论基础则是计算机视觉和光学物理。

2. 数字音频篡改取证方法

针对模拟音频篡改检测的研究 40 年前就有了,而针对数字音频篡改检测的研究则刚开始,公开的研究成果较少。在对数字音频格式、篡改软件、音频分析的校验元组进行分析后,可分别在音频波形统计特征、音频附带背景噪声和音频格式附加信息等三方面进行篡改检测。

1) 基于音频波形统计特征的篡改取证方法

"天然"音频信号在频域上具有很弱的高阶相关性,而大多数篡改操作都会引入一定的非线性,从而导致信号高阶相关性增强,使原来在真实人声频域上很弱的统计相关性变为较为显著的高阶统计相关性。据此检测音频文件是否经过篡改。

2) 基于音频附带背景噪声的篡改取证方法

受图像篡改检测方法的启发,可以利用重采样信号的周期性检查音频中所发生的篡改。不过音频信号的插值检测和图像有所不同:第一,音频在短时内有静音存在;第二,即使没有插值过的音频的局部也可能呈现很强的线性相关。这两点使得 EM 算法无法收敛到理想的结果。为此,可以通过引入音频幅度直方图,排除短时静音和增加样本点数,以使图像的重采样检测算法能有效地用于音频信号的篡改检测。

3）基于音频格式附加信息的篡改取证方法

此外，还可利用音频文件的格式信息进行篡改检测。音频文件格式种类繁多，不同格式的数字音频通常都包含一些必要的附加信息，包括日期、作者、编码格式等。对数字音频材料的篡改很有可能会改变这些附加信息，从而留下篡改痕迹。

3. 数字视频篡改取证方法

与数字图像相比，数字视频的获取设备以及编辑软件的普及度较低，相应地，针对数字视频的取证技术也起步较晚。在 MPEG 文件中，I、P、B 帧的编码方式不同，I 帧只依赖于自身信息进行 JPEG 压缩编码，P 帧依赖于前面的 I 帧或 P 帧的运动估计和运动补偿编码，而 B 帧则利用过去、将来或者同时利用过去和将来的 I 帧或 P 帧作运动估计，再按类似于 P 帧的方式进行编码。当受到篡改时，可能发生帧丢失。通过计算 MPEG 视频流中每个 P 帧的运动误差以及全部帧的平均运动误差，观察运动误差中周期性的噪声，可以确定是否发生篡改。

常见的数字视频篡改检测包括两种情形：第一种是针对消除隔行扫描后的视频；第二种是针对隔行扫描的视频。对于第一种情况，由于消除隔行扫描的两种基本算法是场合并和场扩展，如果将这两种算法看成是一种周期性的插值模式，则可利用 EM 算法来检测插值的周期性。当周期性遭到破坏时，可认为视频遭到篡改。对于第二种情况，通过检测一帧内两个场的运动或相邻帧中场的运动情况，可判断有没有发生篡改。在没有篡改过的视频中，运动是相等的；而在篡改过的视频中，两者不同。由于成像传感器以及摄像机内部电路存在非理想性，在成像过程中必然会产生设备噪声，并被添加到每一帧视频中，而来自同一台摄像机拍摄的视频所包含的噪声存在着相关性。借用同种图像篡改检测的思想，检测前可先从参考视频中计算出摄像机的参考模式噪声，再从待检测视频帧中计算出噪声图像，将噪声图像与参考模式噪声作相关性比较，就可确定是否发生篡改，并可标定出篡改的位置。

5.4 数字内容来源取证

5.4.1 数字内容的来源渠道

数字内容来源设备辨识依赖于这样的假设：同一设备所获取的所有多媒体数据均带有该设备的内在特征，这些特征只与成像/录音管道以及该设备独有的硬件元器件有关，与多媒体数据所表达的内容无关。源设备辨识包含几个不同的层面：设备类型、设备品牌、设备型号以及设备个体，其中设备类型可以是照相机、扫描仪、摄像机、手机和录音机等，设备个体指某一台特定设备。

1. 数码相机成像工作原理

数码相机（Digital Camera，DC）是由镜头、电荷耦合器件（CCD）、模/数转换器（A/D）、微处理器（MPU）、内置存储器、液晶显示器（LCD）、可移动存储器（如 PC 等）和接口（如

计算机接口、电视机接口等)等部分组成，通常它们都安装在数码相机的内部，一些专业的数码相机的液晶显示器与相机机身是分离的。数码相机中的工作原理如下：当按下快门时，镜头将光线汇聚到感光器件 CCD 上，CCD 是半导体器件，它代替了普通相机中胶卷的位置，它的功能是把光信号转变为电信号。CCD 器件上有许多光敏单元，它们可以将光线转换成电荷，从而形成对应于景物的电子图像，每一个光敏单元对应图像中的一个像素，像素越多图像越清晰，如果想增加图像的清晰度，就必须增加 CCD 的光敏单元的数量。CCD 本身不能分辨色彩，它仅仅是光电转换器。实现彩色摄影的方法有多种，包括给 CCD 器件表面加以 CFA(Color Filter Array，彩色滤镜阵列)，或者使用分光系统将光线分为红、绿、蓝三色，分别用 3 片 CCD 接收。这样，就得到了对应于拍摄景物的电子图像，但是它还不能马上被送去计算机处理，还需要按照计算机的要求进行从模拟信号到数字信号的转换，ADC(模数转换器)器件用来执行这项工作。接下来 MPU(微处理器)对数字信号进行压缩并转化为特定的图像格式，例如 JPEG 格式。最后，图像文件被存储在内置存储器中。至此，数码相机的主要工作已经完成，剩下要做的是通过 LCD(液晶显示器)查看拍摄到的照片。有一些数码相机为扩大存储容量而使用可移动存储器，如 PC 卡或者软盘。此外，还提供了连接到计算机和电视机的接口。

数码相机要实现测光、运算、曝光、闪光控制、拍摄逻辑控制以及图像的压缩处理等操作，数码相机通过 MPU 实现对各个操作的统一协调控制。数码相机中的存储器用来保存数字图像数据，与胶卷不同的是存储器中的图像数据可以反复记录和删除。数码相机的输出接口主要有计算机通信接口、连接电视机的视频接口和连接打印机的接口。

拍照手机的成像原理与数码相机基本相同，不同之处在于拍照手机采用的感光器件是 CMOS(Complementary Metal Oxide Semiconductor，互补金属氧化物半导体)，后处理过程相对简单，故分辨率通常都较低。

2. 扫描仪成像工作原理

扫描仪(scanner)是一种高精度的光电一体化的高科技产品，它是将各种形式的图像信息输入计算机的重要工具，是继键盘和鼠标之后的第三代计算机输入设备。从最直接的图片、照片、胶片到各类图纸及各类文稿都可以用扫描仪输入到计算机中，进而实现对这些图像的处理、管理、使用、存储和输出等。

扫描仪主要由光学成像部分、机械传动部分和转换电路部分组成，这几部分相互配合，将反映图像特征的光信号转换为计算机可接受的电信号。光学成像部分是扫描仪的关键部分，也就是通常所说的镜组。扫描仪的核心是完成光电转换的光电转换部件，目前大多数扫描仪采用的光电转换部件是 CCD，它可以将照射在其上的光信号转换为对应的电信号。除核心的 CCD 外，其他主要部分有：光学成像部分的光源、光路和镜头。转换电路俗称机器主板，它负责完成一切电路的伺服工作，A/D 转换工作，当然也包括镜组给它的数字信号的处理。机械传动部分包括步进电机、扫描头及导轨等，主要负责主板对步进电机发出指令带动皮带，使镜组按轨道移动完成扫描。

扫描仪工作时，首先由光源将光线照在欲输入的图稿上，产生表示图像特征的反射光(反射稿)或透射光(透射稿)。光学系统采集这些光线，将其聚焦在 CCD 上，由 CCD

将光信号转换为电信号，然后由电路部分对这些信号进行A/D转换及处理，产生对应的数字信号输送给计算机。当机械传动机构在控制电路的控制下，带动装有光学系统和CCD的扫描头与图稿进行相对运动，将图稿全部扫描一遍，一幅完整的图像就输入到计算机中去了。

扫描仪上的CCD通常包含三列，分别用红、绿、蓝色的滤色镜罩住，从而实现彩色扫描。

3. 数码摄像机工作原理

数码摄像机(Digital Video Camera，DV)基本原理简单地说就是光-电-数字信号的转变与传输。即通过感光元件将光信号转变成电流，再将模拟电信号转变成数字信号，由专门的芯片进行处理和过滤后得到的信息还原出来就是我们看到的动态画面了。由于数码摄像机采用了数字电路，因此数码摄像机具有图像质量佳、记录密度高、可靠性好、成本低以及具有完美的录音音质。

DV与DC的区别在于DV主要用于拍摄连续动态的影像，静态分辨率较低，每帧的数据较少，标准PAL制式和NTSC制式的视频信号，如果换算成像素来表示的话，单幅画面的精度都不足30万像素，即使新兴的高清晰电视HDTV，单幅画面也不过200万像素(1920×1080像素)，在拍照方面，DV的效果是无法和DC比的。

4. 数码录音机工作原理

通常，可通过以下两种渠道获得数字音频。第一种就是将现场声源的模拟信号或已存储的模拟声音信号通过某种方法转换成数字音频；第二种就是在数字化设备中创作出数字音频。音频数字化通常需要经过三个阶段，即采样—量化—编码，具体步骤如下。

(1) 将话筒转化过来的模拟信号以某一频率进行离散化的样本采集，这个过程称为采样。

(2) 将采集到的样本电压或电流值进行等级量化处理，这个过程为量化。

(3) 将等级值变换成为对应的二进制信号(0和1)，并进行存储，这个过程称为编码。

通过上述三个环节，连续的模拟音频信号即可转换成离散的数字信号。

要衡量一个数字音频的音质好坏，通常可以参考以下指标：

- 采样频率：采样点间的时间间隔，通常采用的间隔时间越短，音质越好。
- 量化深度：单位电压值和电流值之间的可分等级数。可分等级数越多，音质越好。
- 音频流码率：数字化后单位时间内音频数据的比特容量。流码率越大，音质越好。

数码录音笔通过对模拟信号的采样、编码将模拟信号通过数模转换器转换为数字信号，并进行一定的压缩后进行存储。而数字信号即使经过多次复制，声音信息也不会受到损失，保持原样不变。

5.4.2 数字内容来源取证方法的评价指标

现有的源设备辨识研究成果主要集中在数字图像，其他源设备辨识的研究成果还不多。不同品牌的数码相机通常使用不同的镜头和成像传感器，并且采用不同的数字信号后处理运算，包括去马赛克、伽马矫正、色彩矫正、白平衡、压缩以及存储等。因此，即使

拍摄同一对象,所生成的数字图像不仅在风格上有所不同,在图像质量上也存在细微差异。提取并分析这些差异特征,可实现对图像生成设备的源辨识。衡量图像源辨识指标主要包括检测率、虚警率和ROC曲线。

(1) 检测率(True Positve,TP):数字内容所对应的来源被正确识别的比率。其计算方式是样本中被正确识别来源的数字内容数目与所有数字内容样本的比值。由检测率可以引申出漏检率的概念,即:

$$漏检率=1-检出率$$

(2) 虚警率FP(False Positive,FP):数字内容所对应的来源被错误识别的比率。其计算方式是样本中被错误识别来源的数字内容数目与所有数字内容样本的比值。由虚警率可以引申出错误否定率的概念,即:

$$错误否定率=1-虚警率$$

(3) ROC曲线:在数字内容来源取证中,通常用ROC曲线来描述TP和FP之间的关系,ROC曲线的AUC可表征分类器的性能。AUC越接近1,表示分类效果越好,反之,AUC越接近于0.5,说明分类性能越差。

5.4.3 数字内容来源取证方法

目前,数字内容来源取证方法主要集中在数字图像来源取证和数字视频来源取证方面,在数字音频来源取证方面的研究较少。为此,这里只对常见的数字图像来源取证方法和数字视频来源取证方法进行介绍。

1. 数字图像来源取证方法

1) 基于图像的统计特征取证方法

基于图像的统计特征取证方法主要利用了图像的统计特征,包括彩色图像R(红)、G(绿)、B(蓝)各个通道上的像素均值,彩色通道RB、BG、GR之间的相关性,各通道上像素相邻分布(统计与各个像素的像素值相差在±1之间的像素个数)的质心,三个彩色通道上图像两两之间的能量比,每个通道上图像三级小波变换后各个子带图像小波系数的均值。除了这些与彩色有关的特征外,还可利用不同相机产生不同质量的图像的特点。客观的图像质量测度可分为三类:基于像素值差异的测度(如均方差、差的绝对值的均值等)、基于相关性的测度(如归一化互相关等)以及基于频谱距离的测度(如频谱的相角和幅值差等)。这些特征构成特征向量,作为支持向量机的输入,进行源设备分类。有些方法还利用了相机镜头特有的径向失真,为了降低生产成本,大部分相机安装了球面镜头。不同型号的相机所安装的镜头不同,其径向失真也不同,因此球面镜头本身的径向失真可以作为设备指纹使用。径向失真的数学表达可由无穷级数描述。以一幅图像的中心为原点,取级数的一阶和二阶系数 k_1 和 k_2 描述径向失真的程度。k_1 和 k_2 可单独组成特征向量作为支持向量机的输入,也可和图像的统计特征联合组成特征向量。利用径向失真作为特征进行分类的主要障碍是径向失真具有随焦距变动而改变的特点,这导致同一镜头的 k_1 和 k_2 不恒定。

2）基于成像设备机器指纹的取证方法

基于成像设备机器指纹的取证方法的思想最早由 Fridrich 等人提出。由于材料的缺陷、工艺的不完善以及半导体的电子噪声，任何成像传感器都有其固有的模式噪声。传感器的模式噪声主要由两部分构成：暗电流所引起的固定模式噪声（Fixed Pattern Noise，FPN）和光敏材料的光子响应非均匀性（Photo Response Nonuniformity，PRNU）所引起的模式噪声。FPN 是加性噪声，中高档相机通过减去一个暗帧可以消除 FPN，所以不宜作为设备水印。但 PRNU 模式噪声（下文直接称为模式噪声）主要由半导体晶片的非均匀性和不完美性产生，一般不易消除，故可当做内部水印使用。模式噪声一个重要的性质是其高频分量与所拍摄的场景无关，并在相机的生命期中相对稳定。据此，若将模式噪声看成一个扩频水印，就可借助水印处理中基于相关性的检测手段来作出判断。获得模式噪声的方法很简单，直接将多幅原始图像减去其低通滤波图像所得到的差值图像进行叠加再求平均，但这种方式所提取的模式噪声易受其他噪声的干扰，包括场景（或称背景）噪声、CFA 插值噪声和 JPEG 压缩噪声等。在检测前通常先对模式噪声做些预处理，以便去除不相干的噪声。例如，Alles 等人提出消除 DCT 块效应，而 Fridrich 等人则将原始图像减去其低通滤波图像所得到的图像认为是残差图像，然后根据统计信号估计理论，利用最大似然估计器从中估计出较精确的模式噪声。不过残差图像中场景噪声、CFA 插值噪声、JPEG 压缩量化噪声以及其他各类噪声的综合影响破坏了利用最大似然估计器所要求的高斯白噪声的假设，导致估计和检测不得不在近似满足高斯白噪声假设的各个分块进行，这使得整个算法的计算量较大。Goljan 等还将第二类方法应用到更复杂的场合，分别对剪切和拉伸后的图像以及从扫描仪所获取的图像进行了来源辨识。除了用于源设备辨识，Goljan 的第二类方法也可进一步推广到图像篡改检测，其工作原理是：若在同一图像中检测到不同成像设备所获取的图像局部，则可确定该图像内容遭到篡改。

3）基于成像管道特性的取证方法

基于成像管道特性的取证方法较特殊，它利用了大多数相机必须使用颜色滤波器这个事实。由于装有单片 CCD 或 CMOS 的相机只能通过颜色插值才能获得彩色图像，而不同厂家、甚至不同型号的相机使用不同的插值算法，因此，只要能从测试图像中估计出插值周期，就可推算出所采用的插值算法，从而追溯出源相机。由于插值点的像素值是由邻域像素的值加权求和而来，借助 EM 算法估计插值系数（即加权系数），并输出一个反映当前像素与其相邻像素相似性的二维概率图，然后在此基础上构造相机品牌的分类器，并且进一步分析了常用的 6 种插值算法，并利用主元分析和神经网络估计插值系数。

归纳起来说，第一类方法利用了图像的统计特征，第二类方法利用了成像设备的机器指纹，而第三类方法利用了成像管道的特性。手机、扫描仪和打印机设备源辨识的方法主要借用了第一、二类相机源辨识的思想。

2. 数字视频来源取证方法

视频来源取证是指根据视频采集过程、处理过程遗留的痕迹确定视频捕获设备，甚至设备型号，以追溯视频的来源。对于互联网上篡改伪造视频的非法传播，来源追溯尤

为有意义。甚至,数字视频的合法版权所有者也可以借助来源取证技术进行视频拷贝检测。

1）基于摄像设备内在特性的视频来源取证

标准视频文件,例如 AVI 文件都包含了文件头信息,可以得到捕获设备、采集时间,分辨率和帧率等信息,但是它们容易被修改,不能作为取证的依据。一种可行的方法是提取视频捕获设备内在固有的一些特征。与图像来源取证类似,数字视频的来源辨识依赖于这样的假设:同一设备所获取的视频数据均携带该设备的内在特征,这些特征只与成像管道以及该设备独有的硬件元器件有关,与多媒体数据所表达的内容无关。与真实性取证类似,这类特征包括相机的镜头失真(chromatic aberration)、CCD 的缺陷或者响应不一致引起的传感器模式噪声 PRNU 等。

2）利用视频码流特征进行来源取证

视频来源取证还可以借助输出数据流的统计特征进行。视频编码标准通常只规定了编码的框架、特征工具和解码器比特流的句法结构等,而编码器的实现具有相当大的灵活性。因此,不同的商家采用了不同的速率控制方案后,每帧输出的码流会在码率的分布控制上会有很明显的差异。甚至,不同的运动估计算法,编码器采用不同的匹配准则、搜索路径等,都可能为视频来源取证提供依据。

5.5 数字内容隐密分析取证

5.5.1 隐密分析取证研究概念及系统模型

隐密通信的系统模型可以用图 5-7 表示。在以数字图像作为载体的情况下,隐密通信过程可以简单地描述为秘密信息经过编码,在密钥的控制下被嵌入到原始图像中形成带有秘密信息的隐密图像;该隐密图像通过信道传输到达接收方,接收方利用已知的密钥从隐密图像中提取出秘密信息;隐密分析是利用秘密信息的嵌入可能引起载体数据分布特性或统计特性的改变,分析在信道中获得的可能的载体信息,从而检测、估计并提取出隐藏的秘密信息。

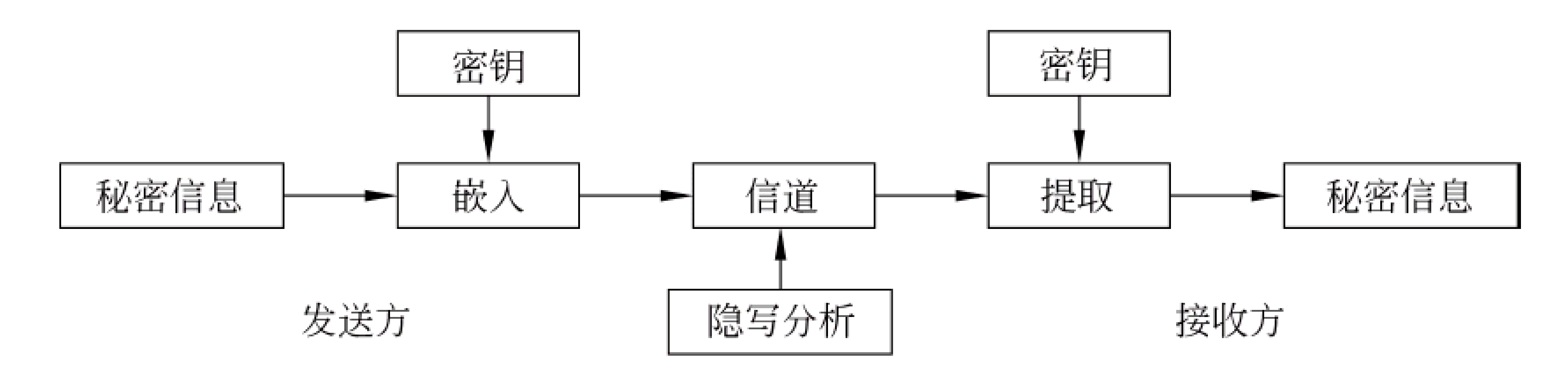


图 5-7 隐秘通信的系统模型

通常所说的隐密分析技术主要指的是证明可疑图像中是否隐藏有秘密信息或者进一步确定秘密信息的嵌入量。而能够作为法律呈堂证据的隐密分析取证技术的目标则是提取并破译出隐藏的秘密信息。所以从一般的意义上来说,隐密分析取证过程由以下几个阶段组成:

(1) 证明可疑图像中是否隐藏有秘密信息。
(2) 确定使用的隐藏方法。
(3) 确定秘密信息的嵌入量和嵌入位置。
(4) 如果嵌入算法使用密钥，寻找嵌入所用的密钥。
(5) 提取出隐藏的秘密信息比特流。
(6) 对秘密信息比特流进行破译解码(密码学范畴)。
隐秘分析取证系统框图可以用图5-8表示。

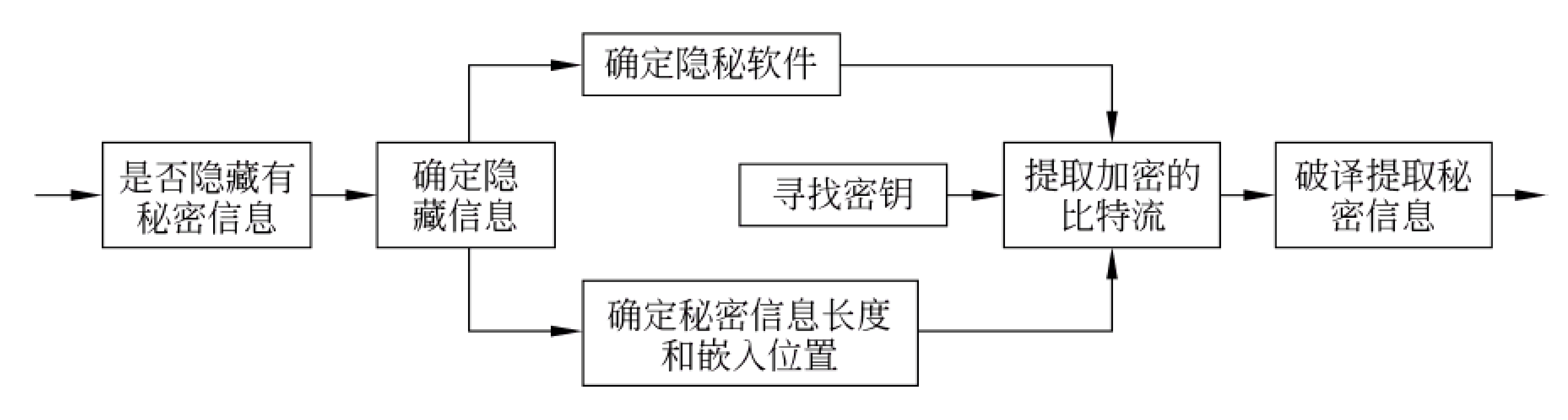


图5-8 隐秘分析取证系统框图

以下分别从一般性的分析方法、针对性的分析方法以及解决隐密分析取证系统中其他问题的分析方法三个角度，对隐密分析取证技术的系统框架进行阐述。

5.5.2 隐密分析取证分类

虽然目前针对性隐密分析方法已经不少，但是对于JPEG格式图像的隐密分析取证的研究还有很多关键问题没有解决，所以，它仍然是隐密分析取证技术研究的主要方向之一。JPEG图像格式比较复杂，对于在DCT域进行的信息隐藏的分析难度较大。因此，在隐密技术的安全性不断提高的情况下，研究JPEG图像的隐密分析取证算法就成为图像隐密分析取证研究中的难点之一。JPEG 2000图像格式是近几年发展起来的图像格式，这种图像格式因其在压缩率、支持无损压缩、支持渐进传输和支持“感兴趣区域”等方面都比JPEG格式更适合于在网络中传输，所以，这种图像格式取代JPEG格式已经成为一种趋势。因此，对于JPEG 2000图像隐密算法的研究已经出现在一些文献中。相应地，有关学者也开始研究针对这种格式图像的隐密分析算法。由于JPEG 2000格式比JPEG格式更加复杂，因此对这种格式图像进行隐密分析也是隐密分析取证技术研究中的难点。

从隐密分析取证系统的角度考虑，现阶段大多数隐密分析研究都集中在检测图像中是否隐藏有秘密信息，而不是提取出秘密信息。如果能进一步解决如何确定隐密所用的方法、软件、密钥这些问题，从而实现对秘密信息的正确提取，就能够使隐密分析进一步达到隐密取证的要求。因此对于隐密分析取证系统中尚未解决的问题，如确定隐藏方法、嵌入软件、嵌入密钥等的研究也是目前亟待发展的方向。

目前，典型的隐密分析手段分为针对性隐密分析算法和通用性隐密分析算法两种。其中，针对性隐密分析算法是针对现有的隐密技术进行分析，因为在分析以前已经知道使用了哪种隐密技术，因此分析起来就较容易一些，这种分析算法的最终目的是力求估

计出携密载体中的含密量。其大体的分析方式如下：

(1) 通过需要分析的掩密技术特征找到一个宏观统计量 S，S 与秘密信息嵌入容量 q 之间存在一种可预测的关系。

(2) 确立 S 与秘密信息嵌入容量 q 之间的关系。

(3) 此关系可能依赖于几个未确定的参数，但可以通过一些极限值(比如原图和最大嵌入量时含密图像的 S 值)来估计这些参数。

(4) 计算待检测图像的统计量 $S(q)$，从而算出秘密信息的嵌入容量 q。通用性分析方法又被称为“盲分析方法”，是通过提取多个通用特征对载体样本集和含密体样本集的训练建立模式分类器，再用此分类器判决待测数据是否含有秘密信息。由于训练样本和待测数据都将映射为多维特征空间中的点，因此此类方法的优劣主要取决于所用统计特征的通用性和分类器的性能。通用性分析方法因为没有针对某一种具体的隐密方式进行分析，所以相对于针对性隐密分析方法要难，其目标是力求发现分析的载体对象中是否含密。例如有基于 BMP 图像、GIF 图像和 JPEG 图像三种不同格式图像的针对特定隐密算法的隐密分析技术。

5.5.3 隐密分析方法的评价指标

对于隐秘分析方法的评价指标主要包括：检出率、漏检率、虚警率、否定率。此外，ROC 曲线也是隐写分析算法评估过程中的一个重要指标。目前比较新颖的评价指标还有检测误差、检测复杂性以及检测容量与检测极限。

1. 检测误差

隐写分析方法的检测误差定义为：

$$P_e = p(e \mid w_0)p(w_0) + p(e \mid w_1)p(w_1) \tag{5-6}$$

其中，$p(w_0)$ 和 $p(w_1)$ 分别表示数字内容中不含隐秘信息和含有隐秘信息的概率；$p(e|w_0)$ 和 $p(e|w_1)$ 分别表示隐写分析算法的虚警率和漏检率。

2. 检测复杂性

对于隐写分析来说，由于要对大量的可疑数字内容进行处理，希望分析算法越简单越好、检测软件运行速度越快越好。因此，隐写分析算法的检测复杂性涉及两个要素：算法复杂性和检出时间。一般来说，算法的复杂性可以由检出时间来反映。因此，在不考虑算法的复杂性的情况下，可以给出检测复杂性的定义。

隐写分析算法单位时间内能分析的数据大小为该算法的检测复杂性，或称作检测复杂度，即：

$$\mathrm{Com} = \frac{1}{n}\sum_{i-1}^{n}\frac{S_i}{T_i} \tag{5-7}$$

式中，Com 表示检测复杂性，T_i 表示检测 n 张图片中第 i 张图片的用时，S_i 为第 i 张图片大小，$S_i=\mathrm{size}(I_i)$。检测复杂性的运算单位为：K b/s，检测复杂性也称为检测速度。

3. 检测容量与检测极限

在分析隐写分析算法检测容量之前，首先给出隐写算法嵌入容量的概念，本书提到的嵌入容量是指从具体隐写算法嵌入的理论角度考虑，所能嵌入的最大信息容量。例如，在分辨率为 256×256 的 lena.jpeg 中用 JPHIDE 嵌入，允许嵌入的最大信息量为 7KB，那么该图片针对 JPHIDE 的嵌入容量则为 7KB。

隐写算法的理论嵌入率的定义为：对于某一隐写算法，将隐写算法嵌入数字内容中的信息量与数字内容对该算法的嵌入容量之比称为该隐写算法的理论嵌入率。

从定义可以看出，理论最大嵌入率的值应该为 100%，即嵌入信息量为最大值时的理论嵌入率。定义理论最大嵌入率的目的是为了方便确定检测上限。

图像隐写分析算法检测容量的定义应该满足以下两个条件：

(1) 检测容量范围内任意一点的检出率 P_i 应该不低于某一给定的值 θ。

(2) 检测容量 C_d 是隐写算法允许的所有嵌入率 R 的子集，并且是满足上一条件的最大子集。

考虑到上述条件，给出隐写分析算法检测容量定义为：对于某一图像隐写分析算法，若 $P_i \geqslant \theta (i \in C_d)$，且 $C_d \subseteq R$，则 C_d 为该图像隐写分析算法的检测容量。一般情况下，秘密信息的嵌入率与检测算法的检出率是正比例关系。把检测容量范围内的最低嵌入率与最高嵌入率称作该隐写分析算法的检测极限，分别为检测下限和检测上限。分别用符号 R_L 和 R_H 表示。这样，检测容量就可以通过式(5-8)进行计算：

$$C_d = R_H - R_L \tag{5-8}$$

5.5.4 常见的隐密分析方法

1. 一般性隐密分析方法

一般性分析方法主要解决的是隐密分析取证系统中的第一个问题，即在没有任何先验知识的前提下确定图像中是否隐藏有秘密信息。一般性分析方法不针对于任何一种嵌入机制，而可以看成是有关信息隐藏安全性定义的具体实践。这种方法假设自然图像可以用一个特征集合来表示。通过对较大图像库中的图像计算特征得到自然图像特征向量的分布估计，并利用人工智能或模式识别的方法设计分类器，从特征空间的意义上区分原始载体图像和隐密图像。尽管一般性分析方法的检测正确率普遍没有针对性的分析方法高，但是一般性的分析方法具有较好的适应性。对于新的隐密方法，针对性的分析方法需要根据其隐藏机制重新设计，而一般性分析方法很可能仅需要重新训练分类器即可。

图像质量回归分析法是 Avcibas 等人较早提出的使用训练分类器的方法进行隐密分析的一般性的分析算法，该算法利用的是图像隐藏信息之后必然会引起图像质量下降的特性。Avcibas 等人还提出了基于最低位平面和第二最低位平面之间的二值相似性的方法，但是这种方法只适用于分析修改最低有效位的隐藏算法。在对一般性的隐密分析方法研究中，Farid 提出的高阶统计量方法比较著名，其思想是对图像进行可分离的正交镜

像滤波器(QMFS)分解,选取不同分解级和不同分解方向(水平、垂直和对角)上每一子带系数的均值、方差、偏度和峰度以及系数幅值的最优线性预测误差等统计量作为区分载体图像和隐密图像的分类特征矢量,并用线性或非线性分类器进行分类判决。该方法能对大多数的隐藏算法进行有效检测,但是检测正确率还没有达到理想的程度。Fridrich在研究JPEG图像的兼容性、分析F5和OutGuess算法之后,综合地提出了针对JPEG图像的基于特征的隐密分析方法,实现了对于JPEG格式图像隐藏算法的一般性分析。该方法对于F5、OutGuess和MB1这些隐密方法都具有较好的检测效果。Harmsen提出了使用直方图的特征函数的重心作为特征的分析方法,这种方法可以用来分析加性的信息隐藏机制,如空域的LSB算法和扩频隐藏算法等。宣国荣等根据Farid和Harmsen的思路,提出了基于小波系数特征函数统计特性的隐密分析方法,其思想是提取图像及小波子带的直方图特征函数的重心作为特征,用贝叶斯分类器进行分类判决。

2. 针对性隐密分析方法

针对性的分析方法是针对不同隐密方法的嵌入机制所设计的分析方法。针对性的分析方法主要解决的问题是在已知嵌入方法的前提下,确定图像中所隐藏的秘密信息的长度(即嵌入容量)。Fridrich给出了用于针对性分析隐密方法的通用方法论(详见5.5.2节)。这种方法的优点是检测时没有门限的约束,不用训练,而且可以估计出嵌入信息的长度。

如果按照图像格式进行划分,针对性的分析方法可以分为针对BMP图像和JPEG图像等图像格式的分析方法。目前针对空域LSB替换算法的分析技术比较成熟,尤其是基于统计的分析方法。其中Westfeld和Pfitzmann提出的chi-square检测算法和Fridrich提出的RS算法是最为著名的方法。RS方法利用嵌入信息前后图像数据相关统计特征的差异来分析隐密图像,但是这种方法仅对于LSB替换的检测有效,而不能分析基于±1以及±K策略的隐密图像。孔祥维等通过分析嵌入信息前后图像的空域特性,提出使用统计滤波和复杂度估计的方法来分析空域隐密算法,实验结果表明该算法对软件Stash-It v1.1具有较好的检测效果。Westfeld提出的算法是分析±1的最早的方法之一,该方法利用了秘密信息的嵌入会增加图像数据中颜色对的数量这一特性。Andrew D. Ker在Harmsen提出的直方图特征函数重心特征的基础上,研究了分别用于检测灰度BMP图像和彩色BMP图像的±1嵌入机制的分析方法,首先对掩密图像用均值滤波器进行降采样处理从而估计原始载体图像,其次提出用相邻像素直方图代替普通直方图,最后用掩密图像与估计出的原始图像的直方图特征函数重心的比值作为特征进行分析。Fridrich和T. Holotyak分别提出了用于检测$\pm K$的检测算法。这两种算法都是利用随机过程的理论来估计非自适应$\pm K$算法的嵌入容量。

JPEG图像格式由于色彩逼真、占用存储空间小,已经成为在互联网上传输的主流格式。目前已经出现了许多针对JPEG图像的隐密算法,如JSteg、FS、OutGuess、MB1、MB2等。相应地,也出现了针对以上隐密算法的分析算法。针对JPEG图像的分析技术最初由A. Westfeld和A. Pfitzmann提出,他们使用统计Chi-square检测对Jsteg隐密算法进行分析,通过用Chi-square检测来判定待测图像的量化DCT系数直方图是否和嵌

入过信息的直方图匹配,从而判断该图像是否嵌入了秘密信息。后来发展的一些隐密算法,如 FS、Outguess 等都作了一定程度的改进,能够抵抗 Chi-square 分析。为此 Niels Provos 提出了一种扩展的 Chi-square 检测的方法。同样也是利用了嵌入后的图像的 DCT 系数直方图中相邻的 DCT 频数很相近的原理,不同的是采用了固定 DCT 系数的样本尺寸,通过移动采样的位置对不同位置的 DCT 系数进行检测,求出一系列的检测概率。但是 Chi-square 检测的方法仅给出了待测图像含有秘密信息的可能性大小,也就是待测图像直方图相对于隐密图像直方图的近似程度,并没有对秘密信息的长度进行估计。另外实验情况也证明 Chi-square 检测和广义 Chi-square 检测的准确性很不稳定,因此在实际应用中的局限性很大。张涛等提出了一种快速有效的针对顺序或随机 Jsteg 类隐密算法的分析方法,该算法根据量化后 DCT 系数一阶统计特性进行分析,对顺序 Jsteg 和随机 Jsteg 均有效,能够较为准确地估计出秘密信息的长度。于小亿等也提出了类似的分析方法,利用广义柯西分布来描述量化后 DCT 系数的一阶统计特性。用这个模型估计出的原始图像量化 DCT 系数直方图与真实直方图非常接近,估计效果很好。全伟伟等利用量化表门限表和单通道量化后 DCT 系数分布,提出了一种新的分析 JSteg 隐密算法的隐密分析方法,该方法能够适用于不同来源的图像。

Fridrich 认为如果隐密后的图像没有因为信息嵌入而留下可检测的痕迹,那么该方法就是安全的。也就是说,隐密图像应该与载体图像有相同的统计特性。FS 算法在嵌入信息的过程中会导致收缩现象的出现,使得量化后的 DCT 系数中的 0 显著增加。虽然它基本保持了 DCT 系数的一阶统计特性,但毕竟不是完全保持。Breaking the FS 算法就是利用了这个改变。根据 FS 的算法,Fridrich 推导出相应的估计秘密信息长度的公式,但其中的参数需要知道原始载体图像直方图的信息。因此她提出用校准的方法来估计原始载体图像:解压缩待测图像,用一个 3×3 的模板进行均值滤波,然后剪切 4 列,再用与隐密图像相同的量化矩阵压缩,认为这种方法得到的图像就是对原始载体图像的估计。空间上剪切 4 列可以打破 DCT 的 8×8 结构,做低通滤波可以消除 8×8 的块效应,这样估计到的载体图像,其 DCT 系数的统计特性与实际的载体图像就十分接近了。其中 Fridrich 也给出了针对双重压缩所采取的措施。但是这种方法对图像的来源较为敏感。在 Attacking the outGuess 的方法中,她同样使用了上述校准估计原图的方法,并使用 8×8 像素块边界的不连续性作为统计量。R. Bohme 提出用一阶统计特征对 MB1 进行分析检测,其思想是利用 JPEG 图像 DCT 系数一阶统计特征的局外系数特性,也就是说,虽然 DCT 系数大致是服从含参数的广义柯西分布,但是总有一些个别的 DCT 系数存在,使得其总体分布不是完全的服从模型分布。由于 MB 算法本身的安全性很高,所以目前还没有更多更好的分析检测方法。

3. 解决隐密分析取证系统中其他问题的隐密分析取证方法

前面介绍的一般性和针对性的隐密分析方法只解决了隐密分析取证系统中的两个问题,即图像中是否隐藏有秘密信息以及如何确定隐藏秘密信息的容量。然而,要达到隐密分析取证的最终目的还必须解决其他的问题:确定隐藏方法或软件,寻找嵌入密钥,提取秘密信息。

在确定一幅图像中隐藏有秘密信息之后，还要分析所用的隐密方法，即解决隐密分析取证系统中的第二个问题，才有可能根据隐密算法提取出秘密信息。Fridrich 等人根据分析 FS、Outguess、MB 所使用的图像校准方法和图像统计特征提出了用多类分类器分析隐密图像中所使用的隐密算法或隐密软件，包括 FS、Outguess、Steghide、JP Hide&seek 和 MB，但是这些方法目前局限于对 JPEG 格式图像的隐密算法进行分类分析。在分析基于密钥的隐密算法中，Fridrich 等人提出了寻找 JPEG 图像隐密算法嵌入秘密信息所用密钥的方法，之后又提出了相应的用于寻找空域隐密算法的嵌入密钥的方法，即解决隐密分析取证系统中的第四个问题：如果嵌入算法使用密钥，还需分析隐密所用的密钥。在获得隐密方法所用的密钥，并已知隐密算法或隐密软件的基础上，就有可能提取出图像中所隐藏的秘密信息。

思　考　题

5.1　什么是数字取证？电子证据有什么特点？
5.2　在各类存储介质中有哪些数据可以作为证据？
5.3　数字取证原则有哪些？写出数字取证过程。
5.4　网络取证有什么特点？请画出网络取证模型。
5.5　写出 IDS 取证的具体步骤。
5.6　利用蜜阱技术进行取证分析时，一般遵循哪些原则？
5.7　模糊专家取证系统包含哪几个组件？
5.8　归纳总结你知道的数字取证工具。

参 考 文 献

[1] Kruse W, Heiser J. Computer Forensics: Incident Response Essentials. Addison Wesley, 2001.

[2] Reith M, Carr C. An examination of digital forensics models. International Journal of Digital Evidence, 2002, 1(3): 1-12.

[3] United States National Institute of Justice Technical Working Group for Electronic Crime Scene Investigation. Electronic Crime Scene Investigation: A Guide for First Responders, July 2001.

[4] 李炳龙，王清贤，罗军勇，等. 文档碎片分类模型及其关键问题. 哈尔滨工业大学学报，2006，38: 834-839.

[5] Li Binglong, Wang Qingxian, Luo Junyong. Forensic Analysis of Document Fragment Based On SVM. 2006 International Conference on Intelligent Information Hiding and Multimedia. 2006: 236-239.

[6] Ianeva T, Vries de A, Rohrig H. Detecting cartoons: A case study in automatic video-genre classification. In IEEE Int 'l Conf. Multimedia and Expo, 2003, 1: 449-452.

[7] Farid H. Image forgery detection. IEEE Signal Processing Magazine, 2009, 26(2): 16-25.

[8] Fridrich J, Soukal D, Lukáš J. Detection of copy-move forgery in digital images. In Proceedings of DFRWS, 2003.

[9] Popescu A C, Farid H. Exposing digital forgeries by detecting duplicated image regions, TR20042515, NH: Dartmouth College, Department of Computer Science, 2004.
[10] Mahdian B, Saic S. Blind authentication using periodic properties of interpolation. IEEE Transactions on Information Forensics and Security, 2008, 3(3): 529-538.
[11] Popescu A C, Farid H. Exposing digital forgeries by detecting traces of resampling. IEEE Transactions on Signal Processing, 2005, 53 (2): 758-767.
[12] 张震,康吉全,平西建,等. 用统计特征量实现的图像拼接盲检测. 计算机应用,2008,28(12): 3108-3111.
[13] Bayram S, I. Avcibas, B. Sankur, et al. Image manipulation detection. Journal of Electronic Imaging, 2006, 15(4): 1-17.
[14] Popescu A C, , Farid H. Statistical tools for digital forensics. Proceedings of International Workshop on Information Hiding. Toronto, Canada: IEEE Signal Processing Society, 2004: 128-147.
[15] Lukas J, Fridrich J. Estimation of p rimary quantization matrix in double compressed JPEG images. http://www. ws. binghamton. edu/fridrich/Research/Doublecomp ression. pdf.
[16] 李晟,张新鹏. 利用 JPEG 压缩特性的合成图像检测. 应用科学学报,2008,26 (3):281-287.
[17] Johnson M, Farid H. Exposing digital forgeries through chromatic aberration. Proceedings ofACMMultimedia and SecurityWorkshop. Geneva, Switzerland, 2006: 48-55.
[18] 王波,孙璐璐,孔祥维,等. 图像伪造中模糊操作的异常色调率取证技术. 电子学报,2006, 34 (12): 2451-2454.
[19] Popescu A C, Farid H. Exposing digital forgeries in color filter array interpolated images. IEEE Transactions on Signal Processing, 2005, 53 (12): 3948-3959.
[20] 张雯,李学明. 改进的基于颜色滤波阵列特性的篡改检测. 计算机工程与应用,2009,45(6): 176-179.
[21] Hsu F, Chang S F. Image splicing detection using camera response function consistency and automatic segmentation. Proceedings of 2007 IEEE International Conference onMultimedia and Expo. Beijing: IEEE Computer Society, 2007: 28-31.
[22] Lin Z, Wang R, X. Tang, et al. Detecting doctored images using camera response normality and consistency. Proceedingsof 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Washington, DC: IEEE Computer Society, 2005: 1087-1092.
[23] Gou H, Swaminath A, Wu M. Noise features for image tampering detection and steganalysis. Proceedings of 2007 IEEE International Conference on Image Processing, San Antonio, IEEE Signal Processing Society, 2007: 97-100.
[24] Johnson M, Farid H. Exposing digital forgeries in comp lexlighting environments. IEEE Transactions on Information Forensics and Security, 2007, 2(3): 450-461.
[25] Johnson M, Farid H. Detecting photographic composite of people. Proceedings of the 6th International Workshop on Digital Watermarking. Berlin: Sp ringer-Verlag, 2007: 19-33.
[26] Maher R C. Audio forensic examination. IEEE Signal Processing Magazine, 2009, 26(2): 84-94.
[27] 高阳. 数字音频材料的真实性检测. 上海:上海交通大学硕士学位论文,2008.
[28] 高阳,黄征,徐彻,等. 基于高阶频谱分析的音频篡改鉴定. 信息安全与通信保密,2008(2): 94-96.
[29] 姚秋明,柴佩琪,宣国荣等. 基于期望最大化算法的音频取证中的篡改检测. 计算机应用,2006,

26(11)：2598-2601.

[30] Wang W H,Farid H. Exposing digital forgeries in video by detecting double MPEG compression. Proceedings of the 8th Workshop on Multimedia and Security. New York：ACM Press,2006：37-47.

[31] 胡永健,刘琲贝,贺前华. 数字多媒体取证技术综述. 计算机应用,2010,30(3)：657-662.

[32] 郑二功,平西建. 针对一类 JPEG 图像伪造的被动盲取证. 电子与信息学报,2010,32(2)：394-399.

[33] 陈威兵,杨高波,陈日超,等. 数字视频真实性和来源的被动取证. 通信学报,2011,32(6)：177-183.

[34] 步山岳,张有东.计算机信息安全技术.北京：高等教育出版社,2005.

[35] 蒋平,黄淑华,杨莉莉.数字取证.北京：清华大学出版社,2007.

[36] Jun-Sun Kim, Minsoo Kim, Bong-Narn Noth. A Fuzzy expet system for network forensics[C]. The 2004 International Conference On Computational Science and Its Application (ICCSA 2004), Perugia, Italy, May, 2004.

第 6 章 chapter 6

文本内容安全

本章学习目标

文本信息是数字内容中最为常见的信息载体之一，其表现形式多样。本章将介绍文本信息的基本概念与文本内容的安全技术，具体包括文本内容加密、文本水印及文本隐写分析技术等。

通过本章的学习，应掌握以下内容：

(1) 文本的概念、分类、特点及表示形式。

(2) 自然语言处理技术。

(3) 文本的加密技术。

(4) 文本的数字水印技术。

(5) 文本过滤及文本分类技术。

(6) 文本隐写及分析技术。

6.1 文本内容安全基本概念

文本数据是信息隐藏中非常重要的一类数据。随着全球信息数字化进程的日益加快，我们日常工作生活中直接接触的各种文本载体资源已经成为人们不可或缺的事物。例如，各种通过互联网传输的文本资源、各种格式的公文处理文档等文本数据，以及扫描成文本图像的个人档案、医疗记录、学历证书、专利证件、手写签名、设计图样、馆藏图书、机要文件等。人们通过频繁地利用上述这些文本资源进行交流、沟通、联系和工作，因此有关文本的版权保护、内容验证等安全问题就成为必须考虑的重要事项，本章将针对文本内容安全的各个方面进行介绍。

由于自然文本中可以有许多种语言，而每种语言的自然文本具有不同的特点，要建立能适用于各种语言的自然文本的精确模型是很困难的。现有的模型一般都是针对某一方面的需求或者根据文本的某个特点而建立的，能从某个角度来对文本做比较细致的解析，而难以从所有方面对文本都有精确的分析，所以对于自然语言的处理和分析是很有必要的。

6.1.1 文本数据的概念、分类及表示

1. 文本数据的基本概念

(1) 文本数据不同于图像的像素编码、视频的帧编码及语音的音频编码等,以字符编码为主来表现信息的数据,它以电子文档的形式存储和传播。一个符号,如果具有记录语言的功能,我们便可以将它视为一个文字,文本即是由这些文字符号所组成的一个序列,文本就是文字信息的数字化表示所形成的电子文件。文本数据编码简单、数据量小、传输便捷,可以和传统的印刷方式的文档进行相互转换、打印、扫描、识别等,因而得到了广泛的应用。

在现有的数字多媒体数据中,许多都是文本数据,如 TXT、DOC、PDF、HTML、XML、EML、XLS、PPT、CHM、WPS、ASP、BAT、BAS、PRG、CMD 以及数据库文件等。针对不同的应用范围、不同的表述对象,文本可以具有不同的描述。在网络化时代,文本数据也是互联网络中最常见和使用最多的一种媒体形式。

(2) 文本图像是把文字资料通过图文扫描仪、数码相机等数据采集设备生成的图像,它不是能用机器立即阅读及处理的文字符号编码文件,而是以数字点阵表示的像素为基本单元进行处理、存储的图像文件。文本图像是以文字、表格、图形等文本信息为主要内容特征的静止图像。文本图像的特点是文字的书写形式与文字所表达的内容同样重要,即若仅将其中的文字提取出来变为普通文本格式进行传输,则会失去或部分失去该文本图像所要表达的内容。在实际应用中,人们常常将文字、表格、图形等文本信息记录在纸张上作为信息存储和交流的基本形式。作为一种灰度图像,文本图像可以通过互联网方便地传输,而在传输的过程中,文本图像的编码格式和数据格式与普通的连续色调的灰度图像都完全相同。

纯文本格式,就是没有任何文本修饰的,没有任何粗体、下划线、斜体、图形、符号或特殊字符及特殊打印格式的文本,只保存文本,不保存其格式设置。将所有的分节符、分页符、新行字符转换为段落标记,用 ANSI 字符设置,只有在目标程序无法阅读任何其他有效的文件格式时才选择这种格式。常见的纯文本格式文件的扩展名:TXT、HTM、ASP、BAT、BAS、PRG、CMD 等。

2. 文本数据的分类

由于文本数据的类型比较多,分类方法也多种多样。

(1) 按内容表现形式可以分为格式化数据和非格式化数据。格式化数据中,编码相同的字符可以有不同的外在表现样式,如文字之间可以设置不同的字距、行距,可以有不同的字体、颜色、尺寸等,如 DOC、WPS、PDF 等数据就属于格式化数据。而非格式化数据中,不同的字符只有编码的不同,没有表现形式的不同,如 TXT 就属于非格式化数据。

(2) 按编码方式的不同可以分为 TXT、PDF、DOC、RTF、HTML 等,通常每种文本编辑器都有自己的编码方法。而同一个文本中的数据根据功能的不同又可以分为:消息主体(message),它是文本中的主体内容,所有表达语义的文字对应的编码数据都是属于

这一类;文档标记(markup),它描述文本的逻辑结构和物理属性,如文本的编码和版本标识,格式化文本中的标记字符以及字体、高度、间距等;附件(appendent),如文本中的图像等额外的非字符编码数据,以及注释等。

3. 文本数据的特点

数字化数据表现的信息对感知系统来说,有的是可以感知的,比如一篇文档中的文字、黑白图像中的像素点对人的眼睛来说是可以感觉到的;而有些信息是感知系统感觉不到的,比如真彩色图像中最低比特位所表现的信息则已超出人眼的感觉范围。这些超出感知系统感知范围的数据,对感知系统来说,就属于冗余数据;另一方面,在信息的数字化过程中,这些冗余部分存在着一定的随机性。那么,将这些具有某种随机性的冗余数据替换为其他随机数据,对感知系统来说是无关紧要的。图像、视频、音频等载体中的信息隐藏正是利用这些数据存在冗余数据的特点,在冗余数据中嵌入信息。

由于文本数据不存在编码冗余,改变其中任何一个比特都将使文本发生可以感知的变化。因此在文本中进行隐藏就不同于图像、音频中的信息隐藏,它需要使用特殊的方式来嵌入信息,文本信息隐藏技术就应运而生了。文本信息隐藏技术是研究各种在文本中嵌入信息的方法,以及如何提高隐藏容量,如何提高嵌入信息的安全性,并根据隐藏方法开发实用的隐藏工具的一门技术。

4. 文本数据的表示

为了以数学形式表示文本的语义内容,常采用向量空间模型(Vector Space Model, VSM),即不考虑文本中词(或其他语义单元)的顺序,将文本简化并表示为特征(feature)权重的向量。其中特征是指文本中的某些词或其他表示语义的单元,所有表示向量所在的向量空间称为特征空间。还有一些 VSM 之外的文本表示模型如基于特征概率分布、基于二维视图等模型、将文本理解为信号序列、高阶词统计(High Order Word Statistics)及 NLP(Natural Language Process)等,但在应用上都还存在局限。

布尔模型是向量模型的一种特例,根据特征是否在文档中出现,特征的权值只能取 1 或 0。许多时候,使用二值特征的分类效果并不比考虑特征频率的差。

目前,VSM 模型相关研究仍然集中在以什么语义单元作为特征及特征的权重确定两个问题上。大部分工作仍旧以词(或 n-gram)作为特征,以特征的频率为基础计算权重,如 TF-IDF(term Frequency-inverse document frequency)等。最新的工作则将一些特征对类别的显著性统计量(如 χ^2 等)引入到权重的计算中,使得支持向量机 SVM 及线性方法的分类效果相对使用 TD-IDF 权重有不同程度的提高。

6.1.2 文本字符的编码方式

文本数据主要是以字符编码的形式来表现信息的数据,本节介绍常见的字符编码方式。

(1) ANSI:系统预设的标准文字储存格式。ANSI 是 American National Standards Institute 的缩写。它成立于 1918 年,是一个自愿性的组织,拥有超过 1300 个会员,包括

所有大型的计算机公司。ANSI 专为计算机工业建立标准,它是世界上相当重要的标准。

(2) Unicode:世界上所有主要指令文件的联集,包括商业和个人电脑所使用的公用字集。当采用 Unicode 格式储存文件时,可使用 Unicode 控制字符辅助说明语言的文字覆盖范围,如阿拉伯语、希伯来语。用户在“记事本”中输入含有 Unicode 字符的文字并储存文件时,系统会提示你必须选取“另存为”中的 Unicode 编码,这些字符才不会被遗失。需要提醒大家的是,Windows 2000 中部分字型无法显示所有的 Unicode 字符。如果发现文件中缺少了某些字符,只需将其变更为其他字型即可。

(3) Unicode big-endian:在 Big-endian 处理器(如苹果 Macintosh 电脑)上建立的 Unicode 文件中的文字位元组(存放单位)排列顺序,与在 Intel 处理器上建立的文件的文字位元组排列顺序相反。最重要的位元组拥有最低的地址,且会先储存文字中较大的一端。为使这类计算机的用户能够存取你的文件,可选择 Unicode big-endian 格式。

(4) UTF-8:UTF 意为通用字集转换格式(Universal Character Set Transformation Format),UTF-8 是 Unicode 的 8 位元格式。如果使用只能在同类位元组内支持 8 个位元的重要资料一类的旧式传输媒体,可选择 UTF-8 格式。

6.1.3 自然语言处理

文本信息隐藏与自然语言处理(Natural Language Processing,NLP)有着紧密的联系,本小节将简要介绍与本文研究工作相关的 NLP 基础。

所谓自然语言,指的是人们日常使用的语言,如汉语、英语、日语等,它是相对于人造的计算机语言而言的。从计算机科学的角度看,NLP 的任务是建立一种计算模型,这种计算模型能够像人那样“理解”自然语言。然而,由于自然语言固有的复杂性,人们对自己理解语言的机制也还是不甚了了。说话人可以用不同的话表达同样的意愿,也可以用同一句话表达不同的意思。反过来,对于同一句话,不同的听话人也会有不同的反应。不过,由于语言是信息的载体,关于计算机对自然语言的理解一般可以根据实用的信息处理的观点来进行评判,如果计算机系统实现了人机会话、机器翻译、自动文摘或抑扬顿挫带有感情地朗读文章等语言信息处理功能,则认为计算机具备了一定程度的理解自然语言的能力。目前,自然语言处理的研究成果已在数据库系统设计、大型软件包、人工智能研究、专家系统设计等领域得到了广泛的应用。

1. 自然语言的分布模型

1) 马尔可夫模型与字母分布

马尔可夫模型认为,自然语言中一个符号对先前的符号有某种依赖性,一个符号是由先前一个或更多的符号决定的,即:

$$P(t_n \mid t_{n-1}, t_{n-2}, \cdots, t_1) = P(t_n \mid t_{n-1}, \cdots, t_{n-k}) \qquad (6\text{-}1)$$

式中,t_i 为文本中第 i 个字符。

由于对先前符号的依赖,字符集中的符号不是均匀分布的。例如,在英语中,通过对大量自然文本的统计,各个字母的出现概率如表 6-1 所示。对自然文本来说,其字符出现的概率跟表所列非常接近。这接近程度可以用相似度来描述:

表 6-1 自然英文文本的字母统计频率

字 母	频 率	字 母	频 率	字 母	频 率
E	0.1268	L	0.0394	P	0.0186
T	0.0978	D	0.0389	B	0.0156
A	0.0788	U	0.0280	V	0.0102
O	0.0766	C	0.0268	K	0.0060
I	0.0707	F	0.0256	X	0.0016
N	0.0706	M	0.0244	J	0.0010
S	0.0634	W	0.0214	Q	0.0009
R	0.0594	Y	0.0202	Z	0.0006
H	0.0573	G	0.0187		

设 $f_1(i)$、$f_2(i)$分别为 2 个概率分布，其中，$i=1,2,\cdots,26$(分别对应 A,B,…,Z)，则 $f_1(i)$、$f_2(i)$的相似度定义为：

$$\alpha = 1 - \frac{\sum_{i=1}^{26} |f_1(i) - f_2(i)|}{\sum_{i=1}^{26} |f_1(i) + f_2(i)|} = 1 - \frac{1}{2}\sum_{i=1}^{26} |f_1(i) - f_2(i)| \qquad (6\text{-}2)$$

2 个概率分布越接近，则相似度越大，完全一样时相似度为 1；反之，2 个概率分布差异越大，相似度越小。自然文本的字符分布与上述参考频率分布的相似度是很高的，当对文本词汇进行某些修改后，相似度会发生改变，且修改越大相似度变化也越大。

2) Zipf(齐普夫)分布模型与单词分布

在自然文本中，不同单词的频率分布是有差别的。Zipf 分布模型认为，排在最频繁出现的单词第 i 位的词频是排在第 1 位的词频的 $1/i^\theta$(θ 为阶数)倍。这表明，设 n 个单词的文本，且其中包含 V 个单词的词汇表，排在最频繁出现单词第 i 位的词频值是$n/(i^\theta H_v(\theta))$，其中，$H_v(\theta)$定义如下：

$$H_v(\theta) = \sum_{j=1}^{v} \frac{1}{j^\theta} \qquad (6\text{-}3)$$

因此，文本中单词按频率降序排列时的频率分布如图 6-1 所示。

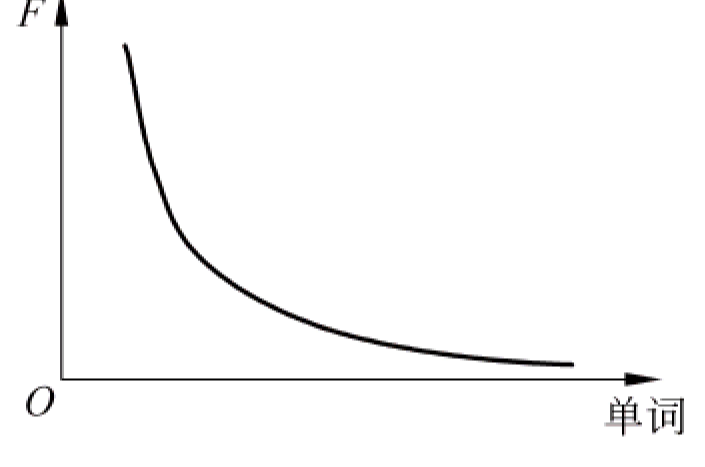


图 6-1 排序的词频分布

统计发现，自然文本中频繁出现(出现次数高于 3 次)的单词占整个文本单词的比例普遍在 15%以上，经常是少数的几个单词占整个文本单词数量的 50%以上，而且文本越长，这个比例也越大。

3) Heaps 定律与单词平均长度及空字符率

Heaps 定律认为，词汇表中单词的长度的增长与文本的大小成对数关系，文本越长就会出现越长的单词。但在整个文本中，因为短词出现的次数也相应地增多，所以单词

的平均长度是一个常数。

根据 Heaps 定律还可得到自然文本中空字符率(空字符的比率)的结论,即空字符率也应接近于一个常数。

空字符不能连续出现两次或多次。当通过空格隐藏法在文本中嵌入空字符后,空字符率将不可避免地变大。

设原文本大小为 T(字节),空字符数量为 s,嵌入空字符数量为 S,则空字符率:

$$R_s = (S+s)/(T+S) = (S/T+s/T)/(1+S/T) = (\mu+b)/(1+\mu) \tag{6-4}$$

其中,μ 是嵌入率;b 是原始文本中的空字符率。对一个给定的文本而言,R_s 随 μ 呈次线性变化,其变化关系如图 6-2 所示。

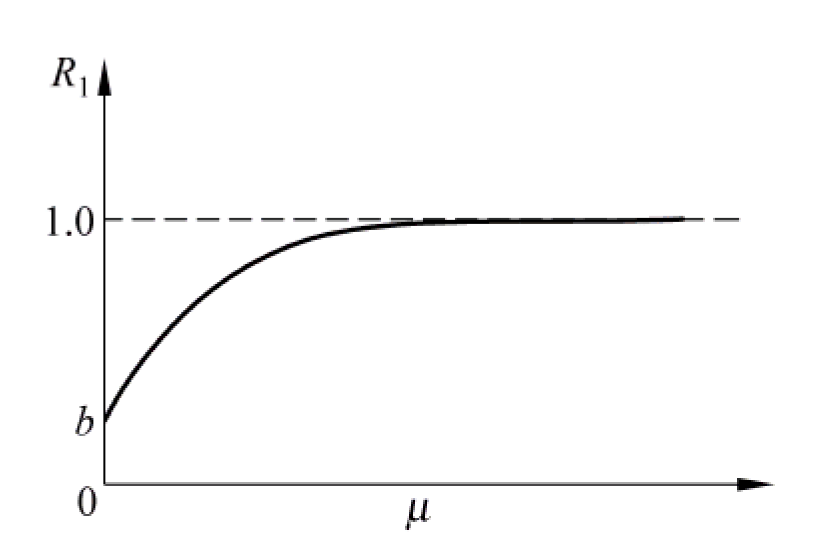


图 6-2 空字符随空字符数量的变化关系

2. 自然语言处理的关键技术

自然语言处理是一个多边缘的交叉学科,除语言学外还涉及计算机科学、数学、统计学、电子工程、心理学、哲学以及生物学等知识领域,它是在各个相关学科的交融和协作中逐渐成长起来的。在历史上,自然语言处理曾经在计算机科学、电子工程、语言学和心理认知语言学等不同的领域分别进行过研究。目前一些关键的自然处理技术有以下几种。

1) 词法分析

所谓的词法分析就是对一个句子或一个词进行切分,使得切分后的词或句子是一个完整的意思,在这一类中最常见的即为文本分词技术。

书面汉语不同于英语、德语、法语等印欧语言,英语、德语、法语在书写时,词与词之间用空格分开,因而词与词之间的界限在书面上是泾渭分明的,而中文文本是按句连写的,例如"计算机科学领域",词间无间隙,通常要将"计算机"划归为一个词,而不是将"计算"单独划分为一个词,因而在中文文本处理中,首先遇到的问题是词的切分问题。按句连写转换为按词连写,词的正确切分是进行中文文本处理的必要条件。1992 年我国出台了《信息处理用现代汉语分词规范》,分词规范的主体结构共分为五大部分:主题内容、适用范围、引用标准、术语概述和具体说明,它将对规范汉语信息处理,对各种信息处理系统的兼容性和语料资源共享起到促进作用。

自 20 世纪 80 年代提出研制汉语自动分词软件以来,已经提出了多种分词方法,如正向最大匹配、逆向最大匹配、有穷多层列举、邻接约束、联想-回溯、词频统计、专家系统、神经元网络等方法。不同的分词方法模拟了人类分词行为的不同侧面,取得了不同的成效,并且已应用在不同用途的中文信息处理系统。

2) 词性标注

词性标注即在给定的句子中判定每个词最合适的词性标记。词性标注的正确与否将会直接影响到后续的句法分析、语义分析,是信息处理的基础。

词性标注的意义在于：

（1）为更高层次的自然语言文本加工提供素材。

（2）为语言学的研究提供翔实的资料。

（3）从加工过的文本中获取词类及频度的词性标注知识。

常用的词性标注模型有 N 元模型、隐马尔科夫模型、最大熵模型、基于决策树的模型等。其中，隐马尔科夫模型是应用较广泛且效果较好的模型之一。

3）句法分析

句法分析就是指对句子中的词语语法功能进行分析，如“妈妈来晚了”，这里“妈妈”是主语，“来”是谓语，“晚了”是补语。句法分析现在主要的应用在于信息处理中，如机器翻译等。它是语块分析思想的一个直接实现，语块分析通过识别出高层次的结构单元来简化句子的描述，从不同的句子中找到语块规律的一条途径是学习一种语法，这种语法能够解释我们所找到的分块结构。这属于语法归纳的范畴。

迄今为止，在句法分析领域中存在很多争议，也许你会发现恰巧有人提出了与你正在努力研究的语法归纳程序偶然产生的相似的句法结构，而且这些也可能已经被当成了句法结构模型的证据。但是，这些找到的结构依赖于学习程序中隐含的归纳偏置。这也指明了另外一个方向，我们需要事先知道模型能够找到什么样的结构，同时应该首先确定我们对句子进行句法分析的目的。

4）语义分析

一个词语“水分”是指物体内所含的水的水分，还是比喻某一情况中夹杂的不真实的成分的水分，在这种情况下就要分析句子的语境，利用语义分析来确定出这个词的意思。

6.1.4 文本内容安全的技术分类

在当今的信息社会中，每天都有大量的信息在传输、交换、存储和处理，在这些日常文档应用、传送、保存过程中既要保证文档数据的安全，又不能影响正常的工作交流，这就关乎文档的安全了。文本安全主要受两方面的影响，一是内部攻击，二是外部攻击。内部攻击是指任何可以访问目标电子文档系统的内部员工都构成威胁。外部攻击是指一些攻击者在强烈的动机驱使下，能够利用多种复杂的策略和技术进行复合攻击，这些攻击者也构成了严重的威胁。所以文本安全工作贯穿着文本管理的每个环节，对电子文件的使用与管理提出了更加严格的安全性要求。它特别强调电子文本的原始性、保密性和完整性，严格防止非授权用户的访问和破坏。

目前常见的文本内容安全技术有：

（1）文本加密技术；

（2）文本的隐写与水印技术；

（3）文本过滤及分类技术；

（4）文本的隐写分析技术。

接下来会分节详细介绍以上每种技术。

6.2 文本内容加密技术

6.2.1 文本内容加密技术的分类

文本加密技术就是保障信息安全的最基本、最核心的技术措施。由第 2 章内容可知,文本加密技术主要是通过对文本数据的加密和数字签名来实现的。其中对数据的加密处理主要是为了防止数据不会被窃听。数据加密的加密方式有两种,一种是传统的对称密钥加密方法,就是加密方用一把密钥对数据进行加密,而解密方用同样一把密钥对数据进行解密。第二种是非对称密钥加密方式,如果使用这种非对称密钥加密算法,它可以保证对发送方和接收方身份的确认。同时数字签名实际上是由生成摘要和生成数字签名两部分构成。其中摘要可以防止文件被篡改,从而保证信息的完整性;而数字签名则是为了保障在商务活动中数据的不可否认性,从而使数据具有法律上的意义。

6.2.2 典型的文本加密方法

1. 基于 Chen-Mobius 变换的文本信息加密法

首先把文本信息变换成数字信息,然后用适当波形对其进行编码,最后加载在 Chen-Mobius 变换$\left(\text{Mobius 变换,即 } t=\dfrac{A\omega+B}{C\omega+D}\right)$加密函数上,通过网络向外传输。在接收端,先通过解密函数进行解密,再通过反变换,恢复原来的文本信息。

文本信息在密码学里面被称为明文,而加载在 Chen-Mobius 变换之后的信息称为密文,通过网络和串口传输的就是密文。系统原理如图 6-3 所示。

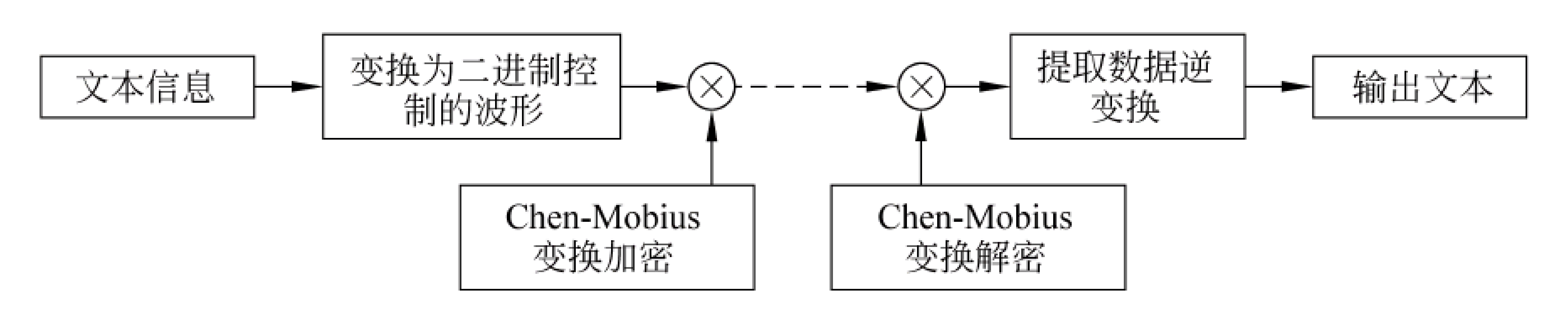


图 6-3 应用了 Chen-Mobius 变换的信息加密系统

应用键控原理,把文本信息变换为二进制的 ASCII 码,而后将其作适当变换,再把二进制的 0、1 用不同的波形编码,然后加载在 Chen-Mobius 变换上形成密文。应用 Chen-Mobius 变换对这些波形加密和解密的实质就是对波形进行调制和解调。

数据处理的原理是:如果用幅值是 1 的一个周期的波形代表 1,幅值是 0.1 的波形代表 0;则幅值 1 的波形的峰峰值为 2,幅值 0.1 的波形的峰峰值为 0.2;通过 for 循环里面设置一个比较空间与 1 进行比较,输出大于等于 1,则输出布尔量 1,否则为布尔量 0。然后八位 0、1 的布尔量,经布尔数组转换成数字的控件变换成数字,再利用数字变换为字符串的控件,变成字符串输出,最后输出文本信息。

2. 字符与汉字的加密方法

Windows 下的字符集采用 Unicode 字符集，它容量大，可置换的范围广。在 Unicode 字符集中，所有字符的内码都占两个字节，因此，如果对 Unicode 字符进行加密运算，需要两个密钥，其取值范围为 0～255，其中一个用于对高字节加密，一个用于对低字节进行加密，这样将某个字符的高、低字节分别加以运算后，生成另外一个 0～255 之间的数，然后再将它们合成为另一个字符，从而置换数据达到数据加密的作用，解密时则相反。例如，密钥 $k_1=68$，$k_2=134$，则字符 A 的低字节为 65，它和 k_1 异或后为 5，A 的高字节为 0，它和 k_2 异或后还是 134，两者合成的字符为“蘅”。再如，“密”的低字节为 198，和 k_1 异或后为 130，它的高字节为 91，和 k_2 异或后为 221，两者合成，则为一个不可见的字符。

3. 序列加密算法

序列加密算法是明文的位串与伪随机数产生器产生的位串经过适当的运算而得到的密文。在序列加密算法中，相同的明文位串可以有不同的密文位串，其结构图如图 6-4 所示。

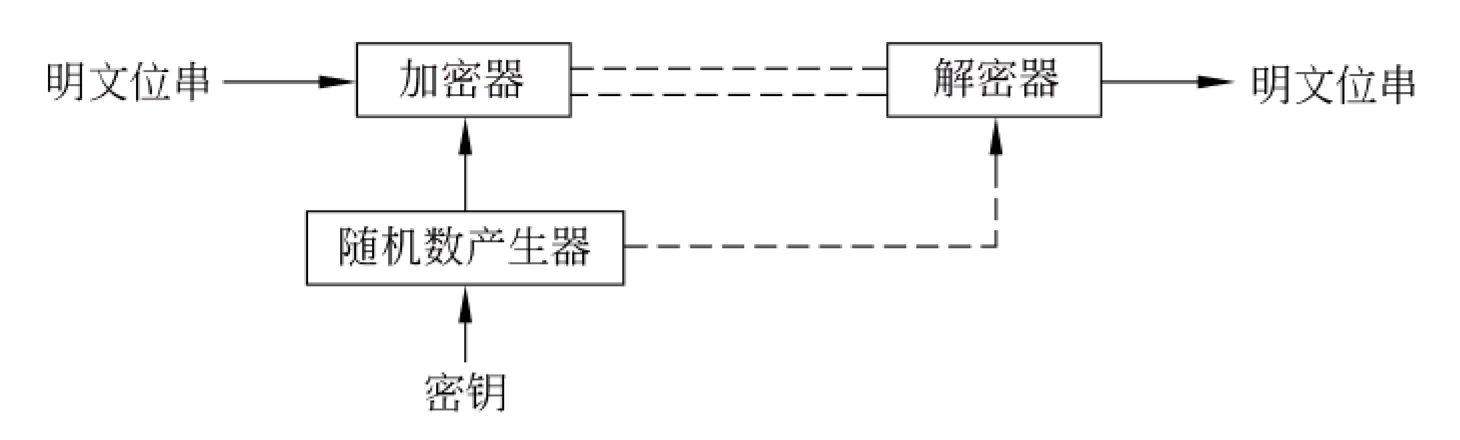


图 6-4 序列加密算法结构图

序列加密算法的安全性在于随机数产生器的密钥位串是否够“乱”，及产生的位串周期是否够长。传统的随机数产生器有：线性反馈移位寄存器、线性同余产生器、非线性随机数产生器及裁剪随机数产生器等。但传统的随机数产生器所产生的伪随机序列加密效果不理想，例如为增强算法的保密性和安全性，利用 Logistic 映射、Henon 映射来作为两个混沌发生器，产生混沌序列对明文进行交替加密。由于 Logistic 映射是一个一维混沌映射，而 Henon 映射是一个二维混沌映射，故利用它们来进行交替加密解决了低维混沌序列保密性不够的缺点。

6.3 文本隐写与文本水印技术

在所有人类的传播媒介中，文字的地位最为重要。可以确定的是，无论传播技术和媒体形式如何变化，文字在传承人类文明和推动社会进步上的作用是其他任何形式的媒体难以比拟的。因此，在文本中隐藏信息及嵌入数字水印进行版权保护、内容认证、操作跟踪具有十分重要的意义，这些技术的重要性是显而易见的。

信息隐藏研究和应用的主要领域有隐写术(steganography)领域和数字水印(digital watermarking)领域。隐写强调如何将秘密信息隐藏在多媒体信息中而不被他人发现，

既隐藏了秘密信息的内容又同时隐藏了秘密信息通信的存在事实。经过隐写处理过的信息与未处理过的信息从表面上看是同样的，混杂在万千的信息之中，使保密通信从“看不懂”转变为“看不见”，容易逃脱攻击者的破解和攻击，如同生物伪装于环境之中免遭攻击一样，此点正是隐写术区别于密码术的根本所在。数字水印指将秘密信息嵌入被保护信息中，用来证明被保护信息的版权、信息完整性、合法使用者等有关内容。数字水印是一种特定的信息，如所有者的名称、标志、签名等，数字水印如同纸币上的水印，传统的水印用来证明纸币上内容的合法性，同理数字水印用以证明数字产品的拥有权、真实性，它是分辨数字产品真伪的一种手段，它隐藏在数字化产品中，人眼看不见，人耳听不着，只有通过数据处理才可识别。数字水印与隐写术不同之处在于数字水印中的载体信息是被保护的信息，它可以是任何一种数字媒体，如数字图像、声音、视频或电子文档，数字水印一般需要具有较强的鲁棒性。数字水印在近年来信息隐藏的研究中占据主要的位置。下面将分别介绍这两种技术。

6.3.1 文本隐写技术

与图像、视频、音频等多媒体中的隐写方法相比，文本中的冗余信息非常有限，所以文本隐写所用的方法与其他几类载体中使用的隐写方法往往截然不同。目前，关于文本隐写方面的研究主要是 J. T. Brassil、N. F. Maxemchuk、S. Low 等提出的格式文本中基于调整行距、字间距、字符特征等来嵌入隐藏信息的方法；M. J. Atallah 等提出的基于同义词替换的嵌入隐藏信息的方法及基于句法和语义的嵌入隐藏信息的方法等。

目前文本隐写算法虽然层出不穷，但是基本上可以归结为以下几类。

1. 基于格式的文本隐写技术

由于文本的冗余度比图像和音频的冗余度要低得多，因此早期的文本隐藏算法设计通常是通过改变原有文本的格式特征来达到隐藏目的。这类隐藏算法可以分为两类：变化间距隐藏算法、变化字体隐藏算法。前者通过改变词汇与词汇之间的距离，行与行之间的间距，甚至段与段的间距达到隐藏目的。例如要隐藏 1 可以将行与行的间距轻微地上移，要隐藏 0 则将行与行的间距轻微地下移来表示。当然，也可以综合用词汇、行和段的间距来隐藏信息。后者是通过改变字体的类型，标点的类型甚至字母的类型来达到隐藏信息的，例如“I am from Beijing”，如果想隐藏 1001 则可以将第 2、3 个词汇的首字母改为斜体即可，“I *a*m *f*rom Beijing”。

1）行间距编码

行间距编码就是在文本的每一页中，每间隔一行轮流地嵌入秘密信息，但嵌入信息的行的相邻两行位置不动，作为参考，需嵌入信息的行根据密文数据的比特流进行轻微的上移和下移。该编码技术具有很强的稳健性，即使经过多次复制，或者页面按某个伸缩因子进行多次缩放，嵌入的秘密信息也可检测出来。

2）字间距编码

通过将文本某一行中的一个单词进行水平移位来嵌入秘密信息。此种方法与行间距码隐藏信息的原理大致相当，都是通过移动间距来实现。相对而言，字间距编码能够

隐藏更多信息,但抗攻击能力较行间距编码要弱。

3)特征编码

通过改变文档中某个字母的某一特殊特征来嵌入标记。特征可以是字体,也可以是b、d、h、k等字中的垂直线,可稍微修改其长度以使一般人不易发觉。相对某种给定字体可以改变其字符高度,如标点信息隐藏方法就是利用中文与英文的标点输入所占用字符宽度的区别来进行信息隐藏的。字体信息隐藏方法是可以利用两种相似的字体,修改文本中一些文字的字体信息来隐藏秘密消息,这些字体被修改后很难被察觉。在不同的字体类型中,有许多字体是非常相似的,如中文的宋体与新宋体之间,楷体与仿宋等,英文中的MingLiu、Times New Roman与Times等都是很相似的,它们在视觉上很难分辨出来,尤其是字符尺寸比较小的时候更是难以区分。如果这些外形相似的不同字体代表不同编码的话,就可以用来进行信息隐藏。格式化文本中,字符可以有不同的颜色。由于颜色用RGB表示,每个色阶用8位共256个级别表示,因而颜色总数可以有1600多万种。而人眼能有效区分的颜色要远远小于这个数字,因而多种相近的颜色人眼根本无法区分。当人眼对文本的颜色的细微区别难以感知时,文本正常显示与使用并不产生任何影响。如果不同的颜色代表不同的编码,那么就可以根据秘密消息比特序列来改变文本的颜色,从而把秘密消息隐藏到文本的颜色当中。此外,通过颜色来隐藏还有一些方式,如把字符颜色设置成背景颜色等。格式化文本中,字符尺寸同样可以用来隐藏消息。当字符尺寸作微小的调整时,人的感知系统就难以感知。如果字符尺寸的不同调整方式与幅度代表不同的编码,那么就可以把秘密消息隐藏在字符的尺寸当中。格式化文本本身具有丰富的字符特征,还可以利用字符的加粗、倾斜、下划线、边框、底纹等特征来隐藏信息。

4)行尾附加空格编码

行尾附加空格编码方法是在每一行的行尾插入空格。每行后最多有几个空格是事先约定好的。如每行后最多有2个空格,编码为1位,4个空格为2位,8个空格为3位,这种方法的好处在于几乎对所有的文本格式均可进行隐藏信息的加载,而且不易觉察。但是这中间也存在着许多缺点,比如通常使用的服务器端软件会提前自动删除文本中的一些多余空格;在对这样的文件进行复制时不会保留所加入的隐藏信息数据。

2. 基于语义的文本隐藏方法

基于语义的文本隐藏算法的基本原理是,将一段正常的语言文字修改为另一段正常的语言文字的过程中将秘密信息隐藏进去,为了防止攻击者发现,算法在修改原文字的过程中使用了同义词替换功能,并在句型的选择、标点的处理、语序重排和错误更正等方面做了许多工作,使得含有隐藏信息的语言文字具有伪自然语言的特征。例如,嵌入文本是“Meeting: 9 o'clock at my home”,掩体文本是“the auto drives fast on a slippery road over the hill.”,隐写文本是“Over the slope the car travels quickly on an ice-covered street.”显然,当攻击者面对隐写文本是很难觉察到隐藏信息的存在的。基于语义的文本信息隐藏算法主要包括同义词替换方法、等价信息替换法。

1）同义词替换方法

同义词替换方法是通过挑出一些词语，用与其意义十分相近的词语进行替换，从而实现秘密消息的隐藏。一对同义词，选用其中一个表示0，另一个表示1。通信双方必须同时拥有同义词表，隐藏信息的容量与同义词表的大小有关。例如用big替换large，人们认为词big是主要词汇，large是次要词汇。由此把文本中这些特定的单词挑选出来构成一个同义词组替换表。需替换的单词表示0，无须替换的单词表示1。在使用这些词时，读者无法适当地认定它们是主要词汇还是次要词汇。但是，当解码时，主要词汇将作为1读出，而次要词汇将作为0读出。这样就可以在文本中隐藏秘密数据。该算法可用于英文或汉语的纯文本中。

2）等价信息替换法

等价信息替换法跟同义词替换相似，是用其他同等属性、具有等价信息量的词汇（短语）来替换文本中的词汇（短语）。与同义词替换中的同义词库类似，等价信息替换中的等价信息主要来源于一个预先建立的事实数据库，库中的信息事先经过编码，隐藏时，根据秘密信息来选择相应的词汇做替换。

3. 基于语法的文本隐写技术

基于语法的文本信息隐藏方法是通过改变句子结构而不显著改变句子意思和语气来达到隐藏目的。这类方法包括句子分拆和组合、主体语前置、宾语前置、主语后置、移动附加语位置、加入删除形式主语、代词替换、主动被动语态变换、加入删除冗余短语等。这类方法的隐蔽性较好，但是受到文本写作风格和内容的影响，在达到较好自然度的同时隐藏容量受到限制，其典型方法有生成文本法。

生成文本法不利用载体文本，而是根据秘密信息直接生成隐秘文本，隐秘文本单纯为了传递秘密消息而生。生成文本法的好坏与字典、构造模型及模板的准确程度有关，它需要一个字典库，即生成文本的“源”，因而也称为“字典法”或者“构造法”。

6.3.2 文本数字水印技术

1. 文本数字水印的概念

文本数字水印技术能提供一种追踪文本被非法复制、发行或伪造的方法。若文本数字水印技术能解决版权问题，传媒业中几乎所有的报刊、杂志、书籍、文件等均可通过网络传播，可节省大量人力、物力和时间，降低成本。另外，大力推行的电子政务方面，也有大量文件在网上流动，如果这类文件被篡改，将会产生严重的后果。

文本水印是用一种无法感知的方法来标记文档，并以此来登记非法分发文档的所有者。如果发现有非法分发的嫌疑，则可以通过检测水印的方法找出文档所有者，它可以很好地解决类似问题。

文本文件没有太多的冗余信息，且在文本文件中嵌入信息极易被阅读者发现，同时一些字处理软件在有意无意间也会破坏原始文件，因而在其中嵌入数字水印比较困难。同时由于文本自身的一些特点，目前用于图像、视频方面的水印嵌入方法大部分不适用

于文档的结构和特性，因而研究文本水印技术已是迫在眉睫。

2. 文本数字水印的载体类型

用于信息隐藏的文本载体主要分以下几种类型：

1）非图像格式的电子文本

包括不具备排版格式的纯文本（如 ASCII 文件、TXT 文件、源程序文件等）和其他具有一定排版格式的文件（如 PDF、RTF、Word、HTML，E-mail 等）。

2）文本图像

包含文本内容的灰度图像或二值图像。常见的是二值文本图像，其中的内容未经文字和排版识别，如传真、乐谱等。

3）纸质文本

从文本信息隐藏角度来看，这类载体若要实现自动提取，需要先对其进行数字化、文字和排版识别等步骤。

3. 文本数字水印的应用

文本数字水印技术的应用是广泛的，从广义上讲，凡是有文字存在的地方，都是文本数字水印技术可能的应用场合。其可能的应用领域如下：

1）数字文本文件的网络发行

互联网上存在大量需要版权保护的数字文本文件（包括文章、杂志等），向这些数字文件中嵌入文本数字水印以宣示文件的版权信息，并作为打击盗版行为的证据，是一种促进数字文本文件网络发行的有力手段。目前，很多数字图书馆（CNKI、VIP 等）、文学作品专业网站（如起点等）均采用了数字水印技术。

2）数字证件、合同的防伪（内容认证）

结合数据加密技术，将载体文本的内容认证信息（签名信息）作为文本数字水印嵌入到载体文本文件中，从而形成有防篡改功能的各种证件、合同，与其他防伪手段相比，文本数字水印技术有着成本小、使用方便以及不影响证书、合同外观质量等特点。

3）重要文件的安全审计

安全审计属于操作跟踪范畴，在很多应用中，需要对涉及国家秘密信息、商业秘密信息的文件进行非常细致的安全防范管理。这些数字文本文件在传输、使用、输出过程的操作人员、时间、存储设备、输出设备等安全审计信息需要被详细地记录，而将这些信息作为文本数字水印嵌入到原文件中，可以极大地方便使用过程的管理与事后的审计。

此外，在隐蔽通信、文件注释、数据库安全、拷贝控制、交易跟踪等领域，文本数字水印技术也有众多可能的应用。

4. 文本水印的常见算法

文本数字水印是数字水印的一种，是以文本为原始载体的数字水印技术。其设计思想和图像数字水印相似：除了文本的作者或者版权拥有者，其他任何人都不能从中检测出水印信息。但是，在文本中加入水印信息更加困难，原因在于和图像、声音中存在的噪

声数据不同，文本中并不包含用于秘密信息传递的冗余信息。

文本水印研究早期的研究是在文档图像(document image)中嵌入水印，采用的方法和图像水印类似，或者利用结构化文档各自格式上的特点嵌入水印，如基于 Word、PDF、PostScript、HTML、XML 等有关的行移编码、字移编码、特征编码、存储物理和逻辑结构、标记变换等。以上方法只考虑保留文本的视觉形式而不考虑其具体内容，通用性较好，隐藏容量较大，但是安全性较差，不能抵御常规的 OCR(Optical Character Recognition，光学字符识别)和格式变换的攻击，而且不能适用于纯文本，应用上也有很大的限制。

长期以来，由于纯文本中没有数据冗余，没有可供插入标记的可感知空间，有学者认为文本是不能被插入水印的。为了向纯文本中嵌入水印，一些学者试图采用插入拼写字母、词的变换、标点符号，甚至一些错误的内容等方法来实现这个目的。一般认为，美国普渡大学 Mikhail J. Atallah 等于 2000 年最先提出了自然语言文本水印的概念。其实早在 1996 年，Bender 等就提出了利用句法和语义变换对文本进行信息隐藏，可以说是自然语言文本水印有关的最早研究之一。

1）基于文档结构微调的文本水印

(1) 行移编码文本数字水印

文本中存在行间距，行移编码就是利用文本的行间距携带水印信息的一种方法。一般在文本中将某一整行垂直移动来嵌入水印信息，而其相邻的上下两行位置不动，作为提取水印时的参照行，嵌入水印信息的行根据水印数据的比特流进行轻微的上移或者下移。根据经验发现，人眼对 1/300 英寸的垂直位移量不敏感，嵌入水印后的文本变化人眼是无法辨认的。

如果一个文本文件最初的行间距是均匀的，那么提取水印时可以通过分析行间距来进行水印提取，不需要原始文本作为参考，可以实现盲提取。行移编码水印算法具有较强的鲁棒性，能够抵抗一定程度的拷贝，缩放攻击，适用于保护打印文档，但是该算法水印容量较小。

(2) 字移编码文本数字水印

字移编码方法是将文本行中的单词在水平位置上左移或者右移来嵌入水印信息，而其相邻的单词不动，作为提取水印时的参考位置。根据经验发现，人眼对 1/150 英寸的水平位移不敏感。通常格式化的文本使用变化的字间距增强文本的可读性和美观感，故利用字移编码方式嵌入水印信息具有更好的隐蔽性。由于最初文档中单词的间距就不均匀，因此提取水印时需要参考原始文档中的字间距。调整单词间距使不同行的平均字间距表现出正弦曲线的规律，从而将水印信息编码到正弦曲线中，增强了鲁棒性，也实现了盲检测，但是水印容量不高。后来，提出利用不同正弦曲线的正交特性提高了水印容量。此外提出了基于统计的字间距编码方法，借助将相邻的单词分组的思想，在文档中重复嵌入相同的水印信息，该方法具有更强的鲁棒性，但水印容量降低。利用字母间距和词间距的改进算法，增加了水印容量，但对于文件拷贝、打印和扫描攻击的鲁棒性较差。

字移编码较行移编码具有更高的水印容量，但是目前已有的字移编码方法在水印提

取时需要参考原始文档,不能实现盲提取。

(3) 特征编码文本数字水印

特征编码通过改变单词字母的特征来嵌入水印。例如字体、字号、颜色、下划线、笔划的宽度、高度、方向、区域的亮度等。此类方法适用于格式化文档和文本图像文档。基于无法识别而扫描仪能够区分字符亮度的事实,提出基于字符亮度调整的方法嵌入水印。依据字符仅有微小差异的新字符来替换原字符的思想,提出在波斯/阿拉伯文中隐藏水印的方法和修改字符的笔画宽度来嵌入水印的方法。基于字符拓扑结构的文本水印算法,利用人类对语言符号的“模糊”心理认知模型和生理视觉模型,通过适当改变字符的拓扑结构,设计出语义上相同的字符的多种字形,用字符字形映射的不同数学模型代表隐藏信息,特征编码的水印容量很大,但是水印提取有时候需要参考原始文档。

2) 基于文本内容的方法

基于文本内容的方法源于信息隐藏技术。它将一个载体文本看成一系列的意义序列而非文本图像,嵌入过程就是将载体文本转换成具有相同或相近意义的隐秘文本的过程。

(1) 同义词替换技术

利用语言的同义词特性,将一组同义词编为不同的隐藏代码,在文本中根据水印信息将原有的词汇进行恰当的同义词替换,使得这些同义词对应的隐藏编码与水印信息匹配,就实现了文本水印的嵌入。检测时不需要原始文本信息,查询约定的同义词对应的隐藏编码得到了水印信息。

这种方法的优点是文本水印与载体文本能非常紧密地结合在一起,水印的鲁棒性、抗攻击性能好,水印不产生视觉影响,但由于难以为所有的词汇找到恰当的同义词,造成文本可嵌入水印信息的容量相当有限。

(2) 基于句法的文本数字水印算法

在自然语言中有很多语义上等价的句法,如主动句与被动句、顺序句与倒装句等,本方法通过修改句子的不同句法来插入数字水印。通过将预先设定的句法映射为隐藏编码,嵌入水印时,首先分析句子的句法,然后恰当调整句子的结构,使得调整后句子句法映射的隐藏编码与要嵌入的水印信息匹配。检测方法可以是明文检测,也可以是盲检测。

本方法的特点为:文本水印的鲁棒性、抗攻击性能好,水印不产生任何视觉影响,但水印的容量有限。同时,由于当前自动分析句法的计算机技术并不成熟,造成嵌入水印的技术相对困难,而且很多情况下会引起语义的变化。

(3) 基于语义的文本数字水印算法

该方法利用语义学原理,将文本描述为文本语义表达(Text Meaning Representation, TMR)树,并将TMR树的不同结构分别对应不同的隐藏编码。嵌入水印时,通过嫁接、剪枝、等价信息替换的方法修改TMR树,同时也对应修改原载体文本。在外在表现形式上,也涉及同义词替换与句法变换,不同点主要在于隐藏信息的编码方法,同时,本方法可以通过添加冗余信息的方式改变原文本。

这种方法鲁棒性好,与内容紧密联系,但水印容量小,自动嵌入水印后可能引起语义

改变,从而降低原载体文本的阅读质量。

3）基于不可见编码的方法

在主要的字符编码标准中(GB 2312、Unicode 等),存在着多个编码对应的字符是不可见的情况,如 ASCII 编码的 32 和 127、Unicode 码默认定义的 OOAOh,此外还可以将 Unicode 标准自定义区的编码定义为空格,水印的嵌入可以利用编码的冗余性和不可见性来进行。典型的方法有替换法与追加法。

(1) 替换法的原理是将多个不可见编码分别对应不同的隐藏信息编码,在文本中对已有的不可见码进行替换,使得替换后的不可见码对应的隐藏信息编码匹配水印信息。这种方法多用于英文这样的语言中,这些语言的文本中单词与单词之间存在自然的空格,因而嵌入的水印容量大,分布也较为均匀,但不适合中文文本,中文文本中字与字之间没有空格。

(2) 追加法的原理是在文本的空白区追加不可见的代码,水印的嵌入、检测方法可以直接依据空白区的空格有无、数量以及不可见编码的类型来确定。常见的方式是在文本段落的末尾空白处添加各种不可见编码,通过与原文的对比来确定是否嵌入了水印,及根据不可见编码的数量与状态来确定嵌入的隐藏信息。用这种方法嵌入的水印容量较大,适合所有语言的文本。缺点是水印信息分布不均匀,而且水印信息极易被恶意攻击者去除。

4）基于图像水印技术的算法

基于图像文本数字水印是将文本转换为图像的形式(如二值图像)进行保存,然后利用加性和乘性、位平面、统计特征、替换、量化、关系等空域水印算法,以及基于离散余弦变换域(Discrete Cosine Transform,DCT)、小波变换域(Discrete Wavelet Transform,DWT)等变换域数字水印算法进行水印的嵌入。

这类文本水印从本质上属于图像水印,其优缺点由不同的图像水印技术确定。总体上,这类文本水印的最突出的问题是载体文本文件的访问方式需要用图像处理软件进行,这与文本通常是通过字处理软件进行访问的方式不一致。

基于图像分块的数字文本水印算法,将文本当作一种特殊的二值图像即黑白像素,通过图像分块,在每一块中嵌入一定量的水印信息,算法需要考虑传统的二值图像水印算法和文本图像的特殊性。主要思想是通过控制“可翻转”的像素并利用“置乱”操作嵌入大量水印信息,水印的提取不需要原始图像的参与,可用于内容认证和篡改提示。

5）基于特殊格式文件的文本水印算法

(1) HTML 网页文件中的水印嵌入算法

从广义上讲,HTML(Hypertext Markup Languge,超文本标记语言)网页也是一种文本,而且它是互联网上最为流行的文件格式。它有着与普通文本文件不同的特点,利用这些特点可以嵌入文本水印。例如,HTML 文本中存在特殊的标记符号,只有恰当的标记包含的数据才能被浏览器显示,这样在正确的标记对之间插入的隐藏信息会被浏览器忽略,从而不会被显示出来。

这种方法的优势是隐藏信息量大,而且不会给用户带来任何视觉上的影响,但与此同时,恶意攻击者可以直接用文本编辑器打开网页查找与修改信息,水印的抗攻击性能

很弱。

(2) 基于 XML 文档的文本水印

XML(Extensible Markup Language,可扩展标记语言)文档是一种广泛用于 Web 的结构化文档,文档本身只有最基本的逻辑结构,物理结构(即显示内容)可根据需要添加。其中的文本 (text)与标志(tags)一起标识着文档的内容,称为 XML 文档;而表达其逻辑结构的基本元素 (elements)和属性(attributes)则可由 DTD(Document Type Defintion)定义。

基于 XML 文档的文本数字水印,就是指在保持 DTD 的约定及文档的应用能力不变的情况下,通过改变 XML 文档的逻辑结构来嵌入秘密信息。这一思想是由日本的 Shingo Inoue、KyokoMak ino 等提出的。图 6-5 给出了这一思想的基本模型。

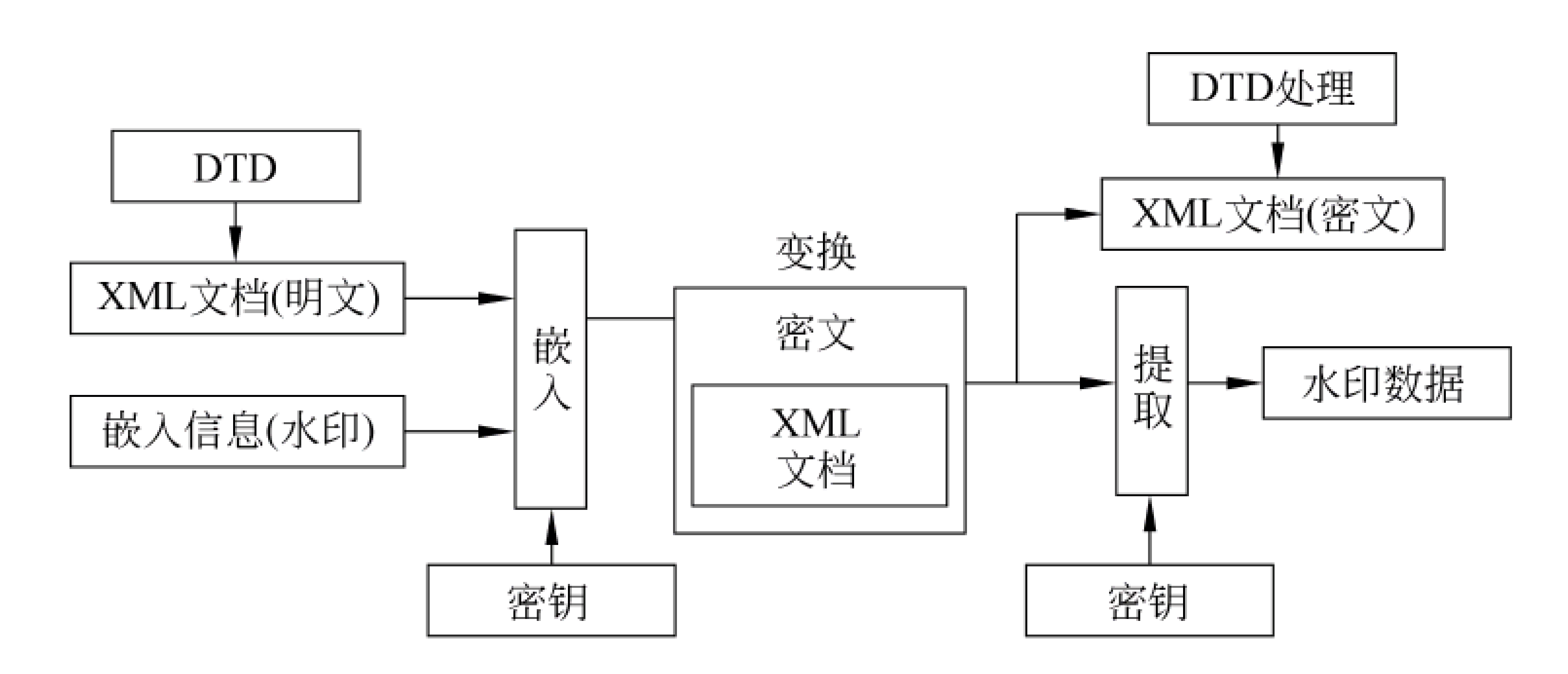


图 6-5 基于 XML 文档的文本水印

在 XML 的应用中,意义相同而逻辑结构不同的文档常常会一起处理。此时只要定义灵活的 DTD,就允许不同逻辑结构的文档存在了。如果相同的处理结果能够来自于具有不同逻辑结构的 XML 文档,那么就可隐藏数字水印于 XML 文档之中。

改变 XML 文档逻辑结构的方式有多种:①变更同名元素的顺序。②置换不同元素的顺序。③使用同义的元素。④使用包含其他元素的元素。⑤使用无意义的空元素。下面以第一种方式为例,详细说明基于 XML 文档的文本数字水印。

在 XML 文档中,当同名元素重复出现时,通过变更这些元素的顺序可嵌入水印。不管这些元素以何种次序书写,在大多数 XML 应用中它们都能被识别为相同意义。

在下面的示例中,元素的字母等级由其内容确定。于是可由元素的等级和嵌入水印的密钥来改变两个元素的顺序,从而嵌入水印。

例 6-1

```
Stego- key(表示嵌入位)
Rank_high_rank_low 0
Rank_low_rank_high 1
Stego- text
< last_name> smith< /last_name>
< last_name> Brown< /last_name>
< last_name> Anderson< /last_name>
< last_name> Woods< /last_name>
```

Embedded data

用这种方法加入的水印不可见性好，不容易去掉，鲁棒性较强。但其局限性也是显而易见的，即它受限于特定的文档结构，且嵌入水印的容量很有限。

(3) PDF 等文件格式中的水印技术

PDF、CAJ 等格式文件自身有固定的格式定义，并与操作系统无关，不需要操作系统提供字体文件来显示字符的字形。在这类文件中嵌入水印的算法并无任何特别之处，可采用前述的各种方法。

由于文本水印的嵌入、检测实现方法有其自身的特点，从而造成了不同格式中的文本水印有不同特点。例如，由于 PDF 文件可与 PS 文件互相转换，而 PS 文件是 Postscript 语言编写的结构化页面描述程序，移动字符(行)间距、修改字符特征、嵌入不可见编码等几乎本节所有的嵌入水印的方式都可以以程序的方式进行自动加载，而且对字符特征信息的获取可以直接从 PS 文件中进行读取，从而实现高效且精确的水印信息检测。这种情况下，利用基于行移、字移的方法进行水印嵌入，可以使得移动的距离很小(等价于增加水印容量)，而且通过精确地读取实际移动的距离，则检测结果精确且高效。此外，这类文件在使用过程中不容易被修改，文本水印的抗攻击能力较强。

6.3.3 典型的文本隐写与水印方法

汉语中除了大量的同义词之外，还存在大量的同音替换现象，如假借、通假、异形词。这些词与同义词的区别在于，它们在发音上、意义上完全相同而只是书写形式不同，而有固定的替代形式。因此没有同义词之间意义上的细微差别和同义词组类的不一致性。利用假借字、通假字、异形词之间的同音替换可以进行信息隐藏。

下面详细介绍一种基于同音词替换的信息隐藏方法。

1. 隐藏原理

(1) 汉语发展过程中出现了大量的假借字、通假字、异形词和异体字，其中很大一部分在现代汉语很少使用或被规范整理淘汰。

例 6-2 利用词对“机伶-机灵”进行同音替换。

原句为：“别看他现在一副傻呆呆的样子，其实他**机伶**得很”。

替换为：“别看他现在一副傻呆呆的样子，其实他**机灵**得很”。

这种隐藏的处理方法与同义词替换法相似，但有两点区别：一是同音替换词表中的各项组成是基本固定的，而同义词表的各项组成对于不同用户会有较大差异；二是对于同义词词对只在某个词性下成立的情况，还需要对同义词进行词性判定。运用这两类方法进行替换时，考虑到汉语句子的词之间没有分割标志，待替换的词与相邻字的组合上可能存在词语组合歧义，所以需先对待处理文本进行分词和切分歧义消除。

(2) 上面提到的同音替换词对在文本中出现的频率有一定的限制，另一种较为通用的替换方式是采用结构助词词对：“的-地”和“的-得”。

“地”作结构助词时用在状语后，表示状语和中心词之间的修饰关系，与“的”可作同

音替换。其中一些典型结构形式为:(副词、形容词)+地+(动词、形容词)。

“得”作结构助词用在动词或形容词后面,连接表示程度或结果的补语,或用在动词和补语中间表示可能时,与“的”可作同音替换四。其中一些典型结构形式为:(动词、形容词)+得+(动词、副词、形容词)。

通过选择“的-地”和“的-得”词对中的前者或后者可分别嵌入1、0信息。

例6-3 利用词对“的-地”进行替换的例子。

原句为:“淘宝购物客户中青年人的数量将会更加快速**的**增长”。

替换为:“淘宝购物客户中青年人的数量将会更加快速**地**增长”。

利用词对“的-得”进行替换的例子。

原句为:“看到爸爸精心准备的生日礼物,她禁不住高兴**的**跳了起来”。

替换为:“看到爸爸精心准备的生日礼物,她禁不住高兴**得**跳了起来”。

采用结构助词词对“的-地”和“的-得”进行隐藏时,需要判定“的、地、得”在句中的词性及相邻中心词的词性,只有符合替换条件的才能进行。每个符合替换条件的结构可嵌入1比特信息。两例句中第一个“的”字位于定语(名词和名词性短语)和中心词(名词)之间,不符合替换条件,故不能用于隐藏信息,应保持为原文状态。词性判定可以人工完成,也可以通过自然语言处理中的句子分词和词性标注算法来实现,利用后者可以实现机器自动隐藏和盲检测。

2. 信息的嵌入

前一种同音替换法的秘密信息嵌入方法和6.3.1节中的同义词替换法相同,不再复述。后一种结构助词词对同音替换法的秘密信息自动嵌入可以按照以下步骤进行:

① 将秘密信息转换成二进制码序列。

② 搜索包含“的”、“地”或“得”的句子。

③ 使用基于规则或基于统计的方法对该句进行自动分词。若分词后“的”、“地”或“得”与其他字一起组成词,或“的”、“地”或“得”处于句首句尾,则这些情况不符合替换条件,只需处理在分词后以单字形式出现在句中的“的”、“地”或“得”。若不存在这种情况,则回到②继续搜索。

④ 对单字形式出现在句中的“的”、“地”或“得”及它们前后相邻的中心词进行词性标注,词性标注有基于规则、统计、机器学习、神经网络或混合的方法。这一步并不需要完成句中所有词的词性标注工作。

⑤ 若词性标注结果存在符合结构“(副词、形容词)+地或的+(动词、形容词)”或“(动词、形容词)+得或的+(动词、副词、形容词)”的情况,则该处可进行嵌入。否则回到②继续搜索。

⑥ 先对原文进行规范化处理,将“(副词、形容词)+地或的+(动词、形容词)”、“(动词、形容词)+得或的+(动词、副词、形容词)”分别规范为“(副词、形容词)+地+(动词、形容词)”、“(动词、形容词)+得+(动词、副词、形容词)”,对比规范化文本嵌入秘密信息。当前待嵌入的秘密信息二进制位为“0”时,保持“地”或“得”不变;为“1”时将“地”或“得”替换成“的”。

⑦ 转步骤②继续嵌入秘密信息的下一个二进制位，直至秘密信息已嵌入完毕或原文已搜索完毕。

以句子“父亲的目的很明确，就是要他努力的学习”为例，说明上述秘密信息嵌入过程。该句的前后两段均包含“的”、“地”或“得”，故先对它们进行分词和词性标注，按照北大版汉语词性标记集的词法分析结果为：

父亲/n 的/u 目的/n 很/d 明确/a，/w 就是/d 要/v 他/r 努力/a 的/u 学习/v。/w

其中 n 表示名词，u 表示助词，d 表示副词，a 表示形容词，w 为标点符号，v 为动词，r 为代词。

第一个和第三个“的”字均以单字形式出现在句中，有待进一步确定是否符合替换条件；第二个“的”字与其他字一起组成词，且出现在句尾，故不符合替换条件。

下面来确定第一个和第三个“的”字是否符合替换条件。前者与其相邻中心词的结构为“n＋u＋n”，即“名词＋助词的＋名词”，不满足替换条件；后者与其相邻中心词的结构为“a＋u＋v”，即“形容词＋助词的＋动词”，符合替换条件。

这样，整句存在一个符合替换条件的结构，故该句可嵌入 1 比特信息。先对原文中符合替换条件的结构进行规范化处理，再根据待嵌入的秘密信息比特是 1 还是 0 分别对规范化文本进行“地(得)-的”替换或是保持不变，最后由图 6-6 所示。

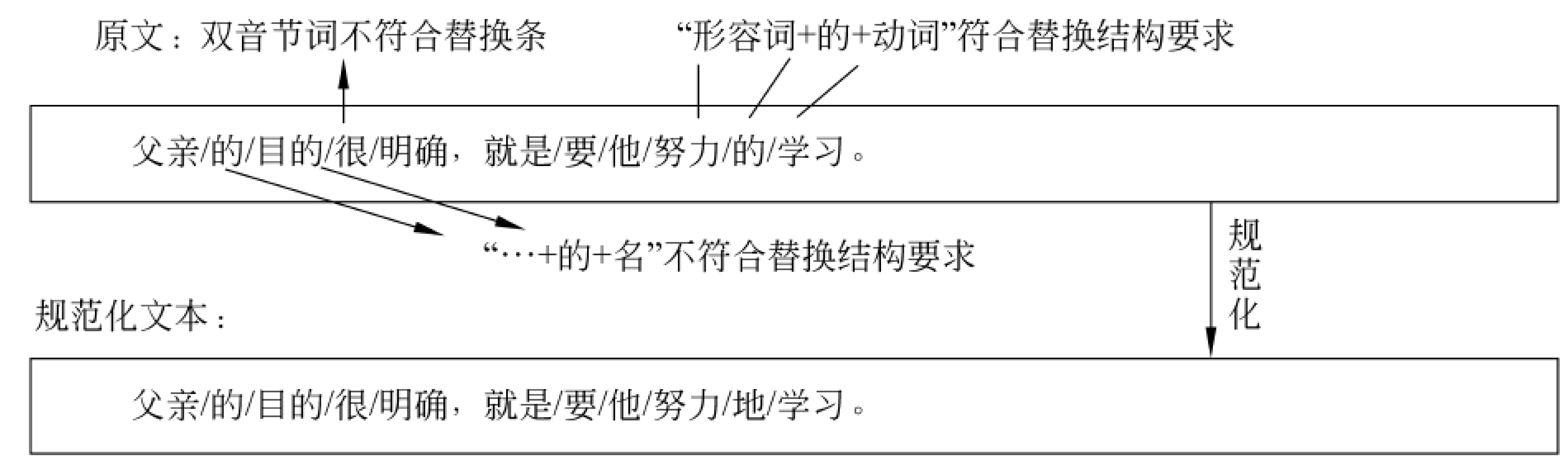


图 6-6 秘密信息嵌入示例

3. 信息的提取

结构助词词对同音替换法的秘密信息自动提取可按以下步骤进行，除对比规范化文本的步骤不同外，它和嵌入方法基本类似。其中所述的原文指已嵌入秘密信息的文本。

① 搜索包含“的”、“地”或“得”的句子。

② 使用基于规则或基于统计的方法对该句进行自动分词。若分词后“的”、“地”或“得”与其他字一起组成词，或“的”、“地”或“得”处于句首句尾，则这些情况不符合替换条件，只需处理在分词后以单字形式出现在句中的“的”、“地”或“得”。若不存在这种情况，则回到①继续搜索。

③ 对单字形式出现在句中的“的”、“地”或“得”及它们前后相邻的中心词进行词性标注，这一步并不需要完成句中所有词的词性标注工作。

④ 若词性标注结果存在符合结构“(副词、形容词)＋地或的＋(动词、形容词)”或

"(动词、形容词)+得或的+(动词、副词、形容词)"的情况,则该处可进行提取。否则回到①继续搜索。

⑤ 先对原文进行规范化处理,将"(副词、形容词)+地或的+(动词、形容词)"、"(动词、形容词)+得或的+(动词、副词、形容词)"分别规范为"(副词、形容词)+地+(动词、形容词)"、"(动词、形容词)+得+(动词、副词、形容词)",将原文与规范化文本进行对比,若"地"或"得"保持不变,则原文嵌入了一个二进制位0;若"的"被规范化为"地"或"得",则原文嵌入了一个二进制位。

⑥ 转步骤①继续提取下一个二进制位,直至原文已搜索完毕。

6.4 文本过滤与分类技术

Internet并不是一个真空环境,它如同一把双刃剑,在给人类造福的同时,也带来了一系列的社会问题。在不同的社会制度、信息环境、文化背景和宗教信仰的影响下,各国用户在这一国际网络空间从事信息发布、传递和使用的过程中,表现出不同的行为规范和道德准则。信息环境的污染主要是由非法信息、有害信息、无用信息造成的。由于世界各国的法律与国情是不一样的,对违法与有害信息的理解与定义也不一样,但有一点是共同的,这些信息对国家安全、社会运行秩序和大多数人的利益构成威胁。这就使得滤除各类有害或无用的信息显得尤为迫切。

当今社会,信息资源已经成为人们竞争的重点。有价值的信息已经成为一种新的财富。大规模文本过滤,作为自然语言处理领域中的一个相当活跃的分支,所研究的内容就是如何准确地表达用户的需求,进而在大规模的信息流中自动地筛选出满足用户特定需求的信息,使人们更有效地利用信息资源。文本过滤技术在帮助人们获取有用信息、滤除无用和有害信息方面起着至关重要的作用,也引起了自然语言处理领域专家学者们的极大关注。

6.4.1 文本过滤技术

1. 文本过滤技术概述

文本过滤技术大致可以分为两类:基于内容的过滤(content-based filtering)和合作过滤(collaborative filtering)。在基于内容的过滤模式中,每个用户假定是相互独立操作的。因此,文本表示仅仅依赖于从文本内容所获取的信息。合作过滤的出发点在于任何人的兴趣都不是孤立的,应处于某个群体当中。在日常生活中,人们接受的信息往往是周围人推荐的结果。因此,根据相同或者相近兴趣的用户对相应文本做出的评注,向其他用户进行推荐。由于不依赖于内容,这种模式不仅适用于文本格式,也可以广泛应用于非文本介质的电子媒介,如MP3、图像、视频等。

文本过滤的任务定义一直在逐渐演化,难度也越来越大。以著名的国际文本检索会议TREC为例:从1997年的TREC-6开始,文本过滤的主要任务逐渐固定下来。以下是从TREC-9至今的文本过滤项目的任务定义:给定一个主题描述(即用户需求),建立

一个能从文本流中自动选择最相关文本的过滤模板(filtering profile)。随着文本流的逐渐进入,过滤系统自动地接受或拒绝文本,并得到文本相关与否的反馈信息,再根据反馈信息自适应地修正过滤模板。

文本过滤项目包含三个子任务:分流、批过滤、自适应过滤。

(1) 分流(routing)子任务。用户需求固定,提供对应于该用户需求的训练文本集中的相关文本,从用户需求构造查询语句来查询测试文本集。

(2) 批过滤(batch filtering)。它和分流子任务很类似。用户需求固定,提供对应于该用户需求的训练文本集中的相关文本,构造过滤系统,对测试文本集中的每一文本作出接受或拒绝的决策;不同的是分流任务要求按相似度从大到小的顺序检索出一批文本,而批过滤则要求将文本分成相关和不相关两类。

(3) 自适应过滤(adaptive filtering)。它要求仅从主题描述出发,不提供或只提供很少的训练文本,逐一判断输入文本流中的文本是否相关。对"接受"的文本,能得到用户的反馈信息,用以自适应地修正过滤模板。而被"拒绝"的文本是不提供反馈信息的。这是最接近真实环境也是最困难的子任务。

2. 文本过滤主要方法

文本过滤的主要方法包括扩展的检索模型和改造的分类模型。下面分别对这两类模型进行介绍。

1) 扩展的检索模型

扩展的检索模型是文本过滤中采用的主要方法,其思路是使用信息检索技术对输入的文档与用户模板之间的相似度进行计算。对得到的相似度使用一个相似度阈值进行过滤任务的决策。相似度高于阈值的文档被认为是相关的。相似度低于阈值的文档被认为与用户模板无关,将被系统滤除。在扩展的检索模型中,一个重要的步骤是通过反馈来提高相似度计算的准确性。同时,阈值的设置和学习也是扩展的检索模型中使用的主要技术。

扩展的检索模型的典型例子是 Okapi 系统,是英国威斯敏斯特大学于 1982 年到 1988 年之间开发的。早期的 Okapi 系统致力于开发基于概率模型的高度交互的检索系统。从 1992 年到 1997 年,英国伦敦城市大学使用基于 Okapi 的过滤系统参加了 TREC 会议(TRECl-TREC6)。从 TREC7 开始,对 Okapi 的改进和应用工作主要由剑桥微软研究中心进行。剑桥微软研究中心还开发了称为 Keen bow 的评价环境。其中一个重要的组成部分是 BSS 系统(Basic Search System)。BSS 系统是基于概率模型的面向集合的检索系统。主要使用倒排索引技术对文本进行检索。该系统致力于权重公式的计算以及对查询的扩展。并提出了著名的权重计算公式 BM25,如下式所示:

$$\sum_{T\in Q} w^{(1)} \frac{(k_1+1)\mathrm{tf}(k_3+1)\mathrm{qtf}}{(k+\mathrm{tf})(k_3+\mathrm{qtf})} \tag{6-5}$$

式中的 Q 是包含项 T 的查询,$w^{(1)}$ 是 Roberso/sparek Jones 权重公式。

$$w^{(1)} = \log \frac{(r+0.5)(R-r+0.5)}{(n-r+0.5)/(N-n-R+r+0.5)} \tag{6-6}$$

式中，N 是文档集中的文档数，n 是包含项 T 的文档数，R 是已知的相关文档数，r 是包含项 T 的相关文档数。

$$K = k_1((1-b) + b \cdot \mathrm{dl}/\mathrm{avdl}) \tag{6-7}$$

k_1、b 和 k_3 都是依赖于查询和文档集的参数。一般情况下，k_1 和 b 分别取 1.2 和 0.75。

k_3 通常取 7 或 1000，tf 是在文档中项 T 出现的频率，qtf 是主题中项 T 的出现频率，dl 是文档长度，avdl 是平均文档长度。

Okapi 系统采用相关反馈的方法进行查询扩展，根据一定的评价指标对待选的项进行排序，根据排序结果选取固定数量的项进行查询扩展。在阈值的调整方面，Okapi 采用 logistic 回归的方法。

2）改造的分类模型

在文本过滤中，判断文本是否符合用户需求可以看作是一个两类(是/否)的分类问题。文本分类的主要方法都可以应用到文本过滤中来。文本过滤中采用的主要分类方法有 Bayes 方法、k 近邻法(kNN)、决策树方法、支持向量机法、神经网络法。

(1) Bayes 法

设训练样本集分为 M 类(文本过滤中 $M=2$)，记为 $C=\{c_1,\cdots,c_i,\cdots,c_M\}$，每类的先验概率为 $P(c_i)$，$i=1,2,\cdots,M$。当样本集非常大时，可以认为 $p(C_i)=C_i$ 类样本数/总样本数。对于一个样本 x，其归于 c_i 类的类条件概率是 $p(c_i|x)$。

则根据 Bayes 定理，可得到 c_i 类的后验概率 $p(c_i|x)$是：

$$p(c_i \mid x) = \frac{p(x \mid c_i) \cdot p(c_i)}{p(x)} \tag{6-8}$$

若 $p(c_i|x)>p(c_j|x)$，$i=1,2,\cdots,M$，$j=1,2,\cdots,M$，则有 $x\in c_i$ 是最大后验概率判决准则，则有：

$$p(c_i \mid x)p(c_i) > p(c_j \mid x)p(c_j), \quad i = 1,2,\cdots,M, j = 1,2,\cdots,M$$

这就是常用到的 Bayes 分类判决准则。经过长期的研究，Bayes 分类方法在理论上论证得比较充分，在应用上也是非常广泛的。从理论上来说，Bayes 分类器具有最优的性能，即所实现的分类错误率或风险在所有的分类器中是最小的，因此该方法常常被用来作为衡量其他分类器设计方法优劣的标准。另外，根据实际情况的不同，以 Bayes 决策为基础，人们还经常使用以下三种分类方法：

① 基于最小风险的 Bayes 决策。如果考虑不同错分情况下有不同的风险，并使错分的风险最小，则此时的 Bayes 决策称为基于最小风险的 Bayes 决策。

② 尼曼-皮尔松分类器。这是一种两类别决策方法，设计原则是在第二类判错的概率保持为常数的情况下，使第一类判错的概率为最小。

③ 最小最大决策。基于最小错误率的 Bayes 决策的一个前提是类别概率 $p(c_i)$是固定的。最小最大决策就是考虑 $p(c_i)$变化的情况下，如何使最大可能的风险为最小，也就是在最差的条件下争取最好的结果，这是一种比较保守的分类方法。

Bayes 方法的薄弱环节在于实际情况下，类别总体的概率分布和各类样本的概率分布函数(或密度函数)常常是不知道的，为了获得它们，就要求样本足够大。

(2) k 邻近法(kNN)

在多数分类问题中,往往不知道类概率密度函数形式,常见的函数形式并不代表实际的真正密度分布。特别是,经典的参数估计大都适用于平滑变化和单峰突出的密度分布,只有一个极大值,而许多实际概率分布却大多是多峰的。在这种情况下就要应用非参数估计统计决策方法,它无须假设类概率密度函数形式为已知条件,而是由训练样本集直接估计类概率密度函数,适用于单峰和多峰情况,包括 kNN 法和 Parzen 等,其中最常用的是 k 近邻法,即 kNN 求法。最初的近邻法由 Cover 和 Hart 于 1968 年提出,是一个理论上比较成熟的方法。该方法的思路非常简单直观:如果一个文本(向量)在特征空间中的 k 个最近邻文本(向量)中的大多数属于某一个类别,则该文本(向量)也属于这个类别。在实际问题中,经常取 $3 \leqslant k \leqslant 7$。与 Bayes 方法比较,kNN 是一个次优方法,因为它使用后验概率的估值作为后验概率。

该方法的优点是简单、准确。理论上,它的错误概率的界限是 Bayes 方法的两倍,但在实际使用上,由于 Bayes 方法的条件比较难于满足,因此,kNN 的效果反而要比 Bayes 方法好一些甚至好很多。该方法的不足是计算量较大,因为对每一个待分类的文本,都要计算它到全体已知样本的距离,才能求得它的 k 个最近邻点。常用的解决方法一是事先对已知样本点进行剪辑,去除对分类作用不大的样本,另一种方法是用空间换时间,事先将所有样本点的两两距离计算出来并存入相应的位置以备检索。前者容易产生新的误差,后者将占用过多的存储空间。

(3) 支持向量机法(SVM)

引入 SVM 的一个重要前提是以往的方法容易产生"过学习"现象。由于 SVM 在训练分类器时充分考虑了小样本情况,因此能够较好地解决这种"过学习"问题。该方法属于研究小样本情况下机器学习规律的统计学习理论范畴,是 1995 年提出的,具有相对优良的性能指标。其核心思想是在分类的误差风险和分类函数的复杂性之间取得一个平衡,从而使分类器对小样本情况具有较好的适应性,克服了"过学习"现象。

支持向量机可以表述为要发现一个超平面 $H(d)=\mathrm{sign}(w \cdot d+b)$,它能将训练集中的数据分开(即训练误差最小),而且有最短的权重向量。支持向量有以下一些特点:

① 它们是各类中与超平面有着最小距离的训练样本。

② 支持向量尽管数量少,但却包含了分类所需的信息。

③ 大部分训练样本不是支持向量,因此移去或减少这些样本对分类器没有影响。

当样本空间非线性可分时,可通过核函数把样本空间映射到一个高维的线性空间,并在新的空间完成点积运算。为了方便计算,通常选择三种特殊形式的核函数作为映射函数:多项式核函数、径向基函数和 sigmoid 函数。

(4) 决策树方法

前述各种方法在数学上表现为构造判决曲面。在实际工作中,构造判决曲面时经常会遇到两种困难:第一,各类之间的界面形状比较特殊,难以用平面、球面或二次曲面方程进行描述。而如果试图构造更复杂的曲面,则工作量会大大增加;第二,各类样本点常常会出现互相混淆的情况,因而难以用一个判决曲面把它们截然分开。

为了克服这些困难,可以采用多级判决的方法,即分成几步进行判别分类。第一步,

利用某一判别方法把待分类的文本分到某几个大组之一。这些大组中可能仍然包括几个不同的类。再对分到各组中的文本进一步判别，然后循此进行，直到把它分到某个确定的类为止。

决策树方法是最常用的分类方法之一。它有以下几个优点：

① 一般来说，由于决策树分类是分成几步来进行的，因此精确度比一次判决要高些。

② 由于决策树分类器是分步进行的，因此每一步的判决规则可以取得简单一些。

③ 每一步不必使用全部的特征，而只使用少数有效特征，这样可以减少每一步的工作量。

④ 分类速度较快。它的不足之处是要在分类器的建立上用较多的时间，还涉及树的结构的确定。另外，对不同的层次上使用的特征和分类方法进行选择以达到最优也是一个比较困难的工作。

(5) 神经网络法

神经网络法的思想是用一系列具有简单计算特性的单元(神经元类型)来组成具有一定数学功能的结构模型，这些单元之间具有广泛的连接(网络结构)，且连接的强度可以根据输入输出数据进行调节(学习算法)。其中常用的神经元模型是阈值函数和sigmoid 函数。

该方法的优点是能够有效地解决很多非线性问题，不足之处是理论上还不够完善，应用时仍有很多因素需要人根据经验来确定，例如初始值的确定和步长的选择，另外还存在着“过学习”问题。目前，人们已经研究出了几十种不同的神经网络，其中在分类上比较常用的有多层感知器、自组织映射和 Hopfield 网。

6.4.2 文本分类技术

文本分类就是将大量文本文档划分为一个或一组类别，使得各个类别代表不同的概念主题。文本分类实际上是一个模式分类任务，所以许多模式分类的算法可以应用到文本分类中。但是，文本分类同时又是模式分类和自然语言处理的一个交叉学科，是和文档的语义紧密相关的，所以与普通的模式分类任务相比有许多独特之处。

20 世纪 90 年代以前，占主导地位的文本分类方法一直是基于知识工程的分类方法，即由专业人员手工进行分类。人工分类非常费时，效率过低。90 年代以来，文本分类技术的研究引起了研究人员的极大兴趣，众多的统计方法和机器学习方法应用于自动文本分类。目前英文自动分类已经取得了丰硕的成果，提出了多种成熟的分类方法，如最近邻分类、贝叶斯分类、决策树方法以及基于支持 SVM、向量空间模型(VSM)、回归模型和神经网络等方法，但对于中文文本的自动分类技术研究尚不尽如人意。目前国内中文文本分类研究主要集中在朴素贝叶斯、向量空间模型和支持向量机等技术上。

文本分类技术的研究主要分为两个阶段：

(1) 基于规则的文本分类阶段

应用知识工程的方法，采用人工方式来构建分类器。大量的领域专家和知识工程师手工编制决策树等推理规则或者专家系统。其优点在于分类的思想容易被人们理解，专家写出的规则可以被大多数人接受，并且对于特定领域和同源的样本分类效果很好。缺

点是需要大量的人来编制规则，开发时间较长而且费用昂贵，而且即使是经验丰富的专家也很难保证规则的一致性和正确性。另外，基于规则的分类系统是针对特定的学科领域和应用环境构建的，不能直接移植于其他的应用系统。这一阶段典型的应用系统是 Carnegie Group 开发的 CONSTRUE 系统，对路透社每天成千上万的稿件进行分类。

(2) 基于机器学习的文本分类阶段

20 世纪 90 年代以后，随着机器学习方法在语音识别等领域取得了很大的进展，越来越多的研究人员开始将机器学习方法引入到文本分类任务中。由于其实现机制简单，构建过程不需要人工干预，并且其分类效果甚至超过了基于规则的系统。因此基于机器学习的文本分类系统很快替代了基于规则的系统，成为研究文本分类的主流方向。

几乎所有重要的机器学习算法都被应用到该任务中，比如最近邻算法(k-Nearest Neighbor，kNN)、贝叶斯、决策树、神经元网络、最大熵模型、LLSF、最小二乘拟和回归模型、支持向量机等。因此目前对文本分类的研究基本上都是基于机器学习的方法。

如图 6-7 所示，构建基于机器学习的文本分类系统的过程大致分为三部分：第一部分是文本表示；第二部分是分类器训练过程，通过对训练数据进行处理得到分类器；第三部分是分类测试过程，将处理过的测试文本输入给分类器，分类器就会给出该文本的类别。

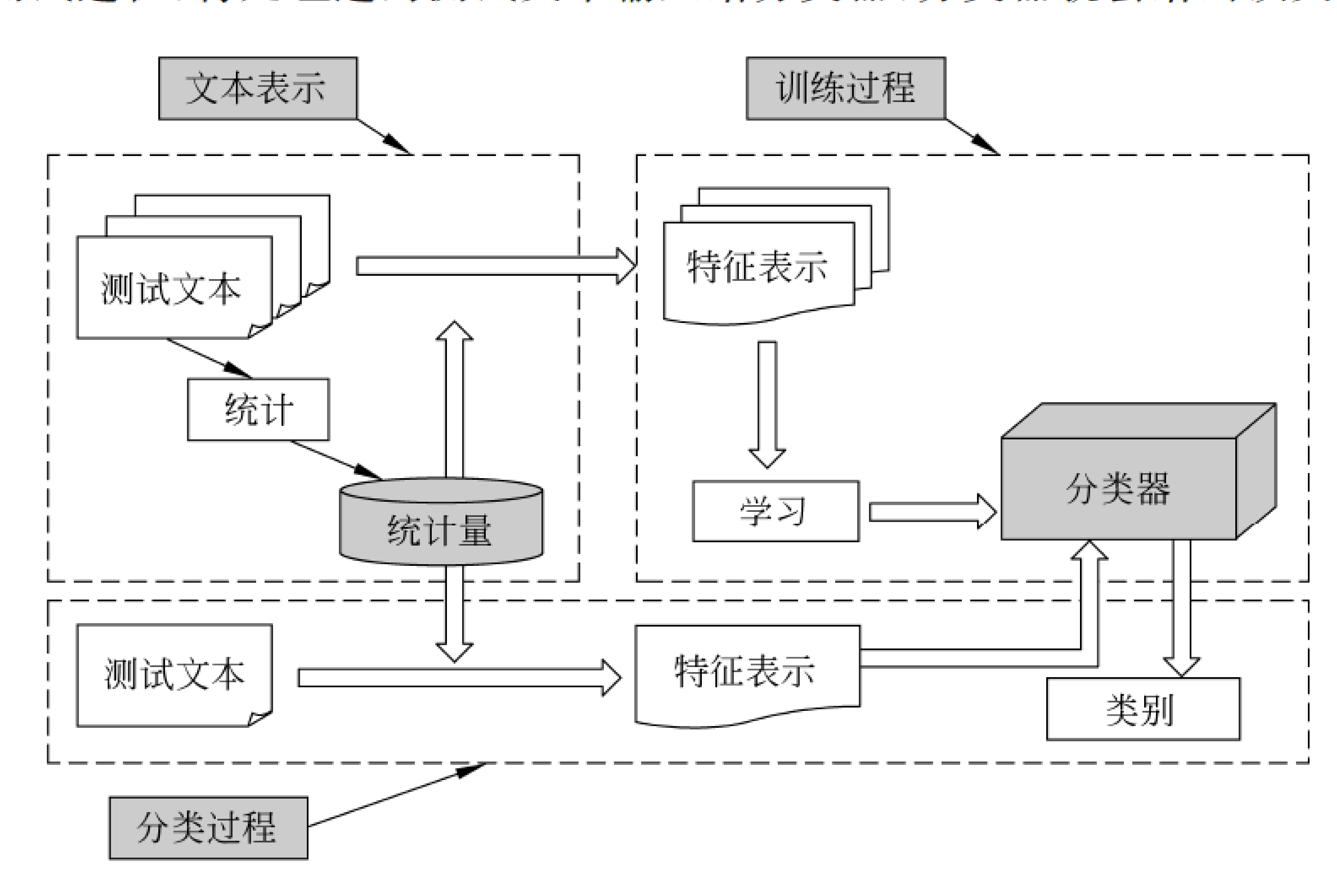


图 6-7 文本分类过程

(1) 文本表示阶段：包括文本预处理、特征降维、权重计算。文本预处理主要是进行去禁用词、词形还原(针对英文文本)、分词(针对中文文本)、词性标注、短语识别等；并且统计词频、文档频率等；经过文本预处理，然后将文本表示成 VSM。在 VSM 中，文本 d 用标引项向量来表示，如下式所示：

$$d=((t_1,w_{1d})\cdots(t_i,w_{id}),\cdots(t_n,w_{nd})) \qquad (6\text{-}9)$$

式中：t_i 为文本 d 中的第 i 个特征，w_{id} 表示 t_i 在文本 d 中的权重，n 为特征集的大小。

特征可以是词、短语、概念、N-gram、词簇等。特征权重计算的方法主要有布尔权重、词频权重、TF-IDF 权重、基于熵概念的权重等。

在文本分类中，向量空间模型是最常用的文本表示方法。在该模型中，文本被表示

为一个向量,向量的每一维对应一个特征。在基于向量空间模型的文本分类系统中,较为常见的是以单词作为特征来表示文本。当把文本表示成特征向量时,其特征数将达到几万甚至是几十万,而且,随着信息量的迅速增多,大规模的文本分类成为人们关注的焦点。因此,在训练分类器之前,首先进行特征降维处理。特征降维的目标是尽量在不影响分类器效果的前提下,去除信息量较少或者不重要的特征。常用的特征降维方式主要有两种:特征选择和特征抽取。特征选择是根据某种准则从初始的特征集中选择比较重要的、类别区分能力较大的特征,其结果是原特征集的子集。目前常用的选取方法有文档频度、互信息、信息增益、解统计、术语强度、特征熵等。特征抽取是将原有的特征进行综合或重组,产生新的特征,即构造从原始特征空间到低维空间的一个变换。抽取的方法主要有潜在语义索引、特征聚类、基于知识库的特征等。

(2) 分类器训练阶段:主要是用各种机器学习的方法对训练数据进行分析,从中学习出各个类别的不同特点,从而生成所需的分类器。目前常用的分类器有 kNN、贝叶斯、决策树、神经元网络、最大熵模型、支持向量机等。

(3) 分类器测试阶段:首先将测试文本进行文本表示,然后利用训练好的分类器对测试文本进行分类。给每个测试文本加上最可能的类别标签。

上面描述的是有监督的分类过程,即训练数据是经过人工标注好的,每个训练文本都有一个类别标签。目前半监督或者无监督的文本自动分类已广泛应用到许多的领域,例如文本的过滤、Web 页面的层次分类、语义消歧、垃圾邮件过滤、主题检测与追踪等文本处理的多个方面。自动文本分类在很多动态的和个性化的信息管理任务中起到重要的作用,比如邮件实时的分类、文件层次分类、垃圾邮件过滤,通过对主题的识别来支持对特定主题的操作等。分类技术既能支持相对静态的分类,例如对于 Medieal subject Headings(Mesh)的分类,雅虎的主题层次分类等,也能支持那些动态的、与个人兴趣相关的分类情况。

文本过滤和文本分类有很大的相似之处。文本分类就是将文本归到若干个类别中。在文本分类过程中,文本的类别可以是预先给定的,也可以是不确定的。前者对应自动分类中的自动归类,后者对应自动分类中的自动聚类。自动归类是分析被分类对象的特征,并与各种类别中对象所具有的共同特征(或一定的分类标准、分类参数)进行比较,然后将对象划归为特征最接近的一类(或最符合标准参数的一类),并赋予相应的分类号。在文本过滤中,判断文本是否符合用户需求可以看作是一个两类(是/否)的分类问题。

6.4.3 典型的文本过滤和分类方法

我们在 6.4.1 节中对文本过滤技术的方法作了较详细的介绍,下面介绍一下文本分类中的一些典型算法。文本分类的方法大部分来自于模式分类,基本上可以分为三大类。

1. 基于统计的方法

1) Naïve Bayes 方法

Naïve Bayes 分类方法(以下简称 NB 法)是一种简单而又非常有效的分类方法。

NB 法的一个前提假设是：在给定的文档类语境下，文档属性是相互独立的。假设 d_i 为一任意文档，它属于文档类 $C=\{c_1, c_2, \cdots, c_j\}$ 中的某一类 c_j。根据 NB 分类法有：

$$p(c_j \mid d_i) = \frac{p(d_i \mid c_j)p(c_j)}{p(d_i)} \tag{6-10}$$

$$p(d_i) = \sum_{j=1}^{k} p(c_j)p(d_i \mid c_j) \tag{6-11}$$

对文本 d_i 进行分类，就是按式(6-11)计算所有文档类在给定 d_i 情况下的概率，概率值最大的那个类就是 d_i 所在的类，即：

$$d_i \in c_j, if\ p(c_j \mid d_i) = \max_{l=1}^{k} p(c_l \mid d_i) \tag{6-12}$$

由式(6-10)和式(6-12)可知，对于给定分类背景和测试文档，用 NB 法分类的关键就是计算 $p(c_j)$和 $p(d_i|c_j)$。计算 $p(c_j)$和 $p(d_i|c_j)$的过程就是建立分类模型的过程。

根据 $p(d_i|c_j)$计算方式的不同，可以将 Naïve Bayes 方法分为最大似然模型(maximum Likelihood model)、多项式模型(multinomial model)、泊松模型(poison model)等。

2) kNN 方法

kNN 法即 k-Nearest Neighbor 分类方法，这是一种稳定而有效的文本分类方法。采用 kNN 方法进行文档分类的过程如下：对于某一给定的测试文档 X，在训练文档集中，通过相似度找到与之最相似的 k 个训练文档。然后根据式(6-13)计算 X 与每个类别的相似度，并按相似度进行排序。还应当设定一个阈值，只有分值超过阈值的类才予以考虑。测试文档属于超过阈值的所有类。

$$\text{score}(X, c_j) = \sum_{X_i \in \text{kNN}} \text{Sim}(X, X_i) y(X_i, c_j) - b_j \tag{6-13}$$

其中，

$$y(X_i, c_j) = \begin{cases} 1, & X_i \in c_j \\ 0, & X_i \notin c_j \end{cases} \tag{6-14}$$

b_j 为阈值，$\text{Sim}(X, X_i)$为文档 X 和 X_i 的相似度，$\text{score}(X, c_j)$为测试文档 X 属于 c_j 类的分值，符号 kNN 表示文档 X 的 k 个最近邻组成的文档集合。

对于某一特定类来说，b_j 是一个有待优化的值。一般 b_j 可以通过一个验证文档集来进行调整。验证文档集是训练文档集的一部分。根据式(6-13)的结果，可以确定测试文档的类别。很显然，对于每一个测试文档，必须求解它和训练文档库中所有文档的相似度。

3) 类中心向量方法

类中心向量方法是一种基于向量空间模型的方法。在分类器的训练阶段，使用训练文档集得到每一个类别所对应的中心向量。在分类阶段，对于某一给定的文档 d，计算文档向量和每个类别中心向量的相似度，然后按相似度进行从大到小排序。相似度最大值所对应的类别，就是文档的所属类别。如果希望文档可以属于多个类别，可以设定一个阈值，文档属于相似度超过阈值的所有类。

常用的获得类别中心向量的方法有以下几种：

(1) Rocchio 方法

Rocchio 方法是一种批处理的学习方法，算法从已存在的向量 $\boldsymbol{w}_1$，及一组训练实例中产生一个新的权重向量 $\boldsymbol{w}$，$\boldsymbol{w}$ 的第 j 个分量 w_j 为：

$$w_j = \alpha w_{1,j} + \beta \frac{\sum_{i \in c} x_{i,j}}{n_c} - \gamma \frac{\sum_{i \notin c} x_{i,j}}{n - n_c} \tag{6-15}$$

式中，n 是训练样本的数目，$C=\{1\leqslant i\leqslant n: y_i=1\}$ 是正类训练样本集，n_c 是正类训练样本数。参数 α、β、γ 则分别用来控制初始权重向量、正例、负例的相对影响，通常设 $\alpha=0$，$\beta=1$，$\gamma=1$。

(2) Windrow-Hoff 算法

LMS 或 Windrow-Hoff 算法是一种在线学习算法，每次使用一个训练样本对旧的类别向量权重进行更新。初始状态下，可以设置 $\boldsymbol{w}_1=(0,\cdots,0)$，当然也可以将其设为其他值。每一步，新的权重向量 $\boldsymbol{w}_{i+1}$ 从旧的权重向量 w_i、使用带标签 y_i 的训练实例 x_i 计算而来，新权重向量的第 j 个分量通过以下公式来计算：

$$w_{i+1,j} = w_{i,j} + 2\eta(w_i + x_i - y_i)x_{ij} \tag{6-16}$$

其中，参数 $\eta>0$，我们将其称为学习率，用来控制权重向量 $\boldsymbol{w}$ 的变化速度，以及每一个新的训练样本对它的影响。WH 可以看作一个梯度下降过程，因为 $2(\boldsymbol{w}\cdot x-y)x$ 恰好是平方差 $(\boldsymbol{w}\cdot x-y)^2$ 的梯度。因此，WH 总是沿着方差减小最快的方向来移动。

(3) EG 算法

EG(Exponentiated-Gradient)算法类似于 WH 算法，每次使用一个训练样本对旧的类别向量进行更新。但是，类别向量的第 i 维限制为非负值，并且进行了归一化。初始时刻，设置权重向量的每一维都相等，即 $\boldsymbol{w}_i=(1/d,1/d,\cdots,1/d)$，$d$ 为特征空间的维数。EG 算法的更新规则如下：

$$w_{i+1,j} = \frac{w_{i,j}\exp(-2\eta(w_i x_i - y_i)x_{i,j})}{\sum_{j-1}^{d} w_{i,j}\exp(-2\eta(w_i x_i - y_i)x_{i,j})} \tag{6-17}$$

4) 回归模型

回归模型中最为典型的就是 LLSF 方法。给定训练文档集和文档类集，LLSF 将其表示为两个矩阵 $\boldsymbol{A}$ 和 $\boldsymbol{B}$。$\boldsymbol{A}$ 代表原始空间，矩阵的第 i 行、第 j 列的元素代表文档 T_i 中的第 j 个特征的权值。$\boldsymbol{B}$ 代表目标空间，矩阵的每一个元素只能取 0 或 1。如果文档 T_i 属于类别 c_i，那么矩阵的第 i 行、第 j 列的元素取值为 1，否则取值为 0。

统计文档中所有词的出现频率，使用 IDF 计算词的权值，可以得到矩阵 $\boldsymbol{A}$；统计每一篇训练文档的所属类别，可以得到目标空间矩阵 $\boldsymbol{B}$。这样文本分类问题就转换为求一个满足条件 $(\boldsymbol{B}_{m\times l})^{\mathrm{T}}=F_{l\times n}\cdot(\boldsymbol{A}_{m\times n})^{\mathrm{T}}$ 的矩阵 $\boldsymbol{F}_{l\times n}$ 的问题，其中 $\boldsymbol{F}_{l\times n}$ 是一个词类别关联矩阵，行代表类别，列代表词。LLSF 就是寻找矩阵 $\boldsymbol{F}_{l\times n}$，使得下式的值最小。

$$\sum_{i=1}^{m} \| \boldsymbol{e}_i \|^2 = \sum_{i=1}^{m} \| \boldsymbol{F}\boldsymbol{a}_i^{\mathrm{T}} - \boldsymbol{b}_i^{\mathrm{T}} \|^2 = \| \boldsymbol{F}\boldsymbol{A}^{\mathrm{T}} - \boldsymbol{B}^{\mathrm{T}} \|^2 \tag{6-18}$$

解决 LLSF 的方法之一是将矩阵 $\boldsymbol{A}$ 进行 SVD(奇异值分解)，

$$\boldsymbol{F} = \boldsymbol{B}^{\mathrm{T}}(\boldsymbol{A}^{-1})^{\mathrm{T}} = \boldsymbol{B}^{\mathrm{T}}\boldsymbol{U}\boldsymbol{S}^{-1}\boldsymbol{V}^{\mathrm{T}} \tag{6-19}$$

给定一篇文档 $\boldsymbol{d}=\{w_1,w_2,\cdots,w_n\}$，可以通过计算 $\boldsymbol{y}=(\mathrm{Fd}^{\mathrm{T}})^{\mathrm{T}}$，将其映射到目标空

间。$\boldsymbol{y}=\{y_1,y_2,\cdots,y_l\}$中每一个 y 的取值为$-1\sim1$，表示文档 $\boldsymbol{d}$ 与每一类的相关度。

5）最大熵模型

最大熵模型(maximum entropy model)的基本原理是拟合所有已知事实，保持对未知事件的未知状态。换言之，就是给定一些事实集，选择一种模型与现有事实一致，对于未知事件尽可能使其分布均匀。它可以对非常广泛的自然语言现象建立概率模型，综合观察到的各种相关或不相关的概率知识，对许多问题的处理结果都到达或超过了其他方法的最好结果。最大熵模型被广泛地应用于自然语言处理中，包括分词、词性标注、词义排歧、短语识别、机器翻译等。

6）支持向量机

支持向量机是在统计学习理论(statistical learning theory)的基础上发展而来的一种机器学习方法，它基于结构风险最小化原理。其基本思想是构造一个超平面作为决策平面，使正负模式之间的间隔最大。支持向量机在解决小样本、非线性及高维模式识别问题中表现出了许多特有的优势，并能够推广应用到函数拟合等其他机器学习问题中。

SVM 已初步表现出很多优于已有方法的性能，并在很多领域得到了成功的应用，如人脸识别、手写体识别、文本分类等。在文本分类方面 SVM 的表现尤为突出，其分类性能几乎超过了现有的所有方法。

2. 人工神经网络

人工神经网络(Artificial Neural Networks，ANN)是对人类大脑系统的一阶特性的一种描述，是一个并行、分布处理结构，它由处理单元及其称为连接的无向信号通道互连而成。具有信息分布存放、运算全局并行、处理的非线性等特点，适用于学习一个复杂的非线性映射，主要应用于语音、视觉、知识处理、辅助决策等方面。根据网络结构和学习算法的不同，人工神经网络分为多层感知器、自组织映射和 Hopfield 网等。

下面以 BP 网络为例来说明人工神经网络在文本分类中的应用。BP 神经网络就是采用 BP(Back Propagation)算法进行训练的多层感知器网络，该网络具有一个输入层，一个输出层和至少一个隐藏(中间)层。一般情况下，选用一个隐藏层就足够了，此时其结构如图 6-8 所示。BP 算法是非循环多级网络的训练算法，其学习过程由正向传播和反向传播组成，输入值经过非线性变换从输入层经隐单元逐层处理，并传向输出层，每一层神经元的状态将影响到下一层神经元状态，如果在输出层不能得到期望的输出，则转入反向传播，通过修改各神经元权值，使误差信号最小。

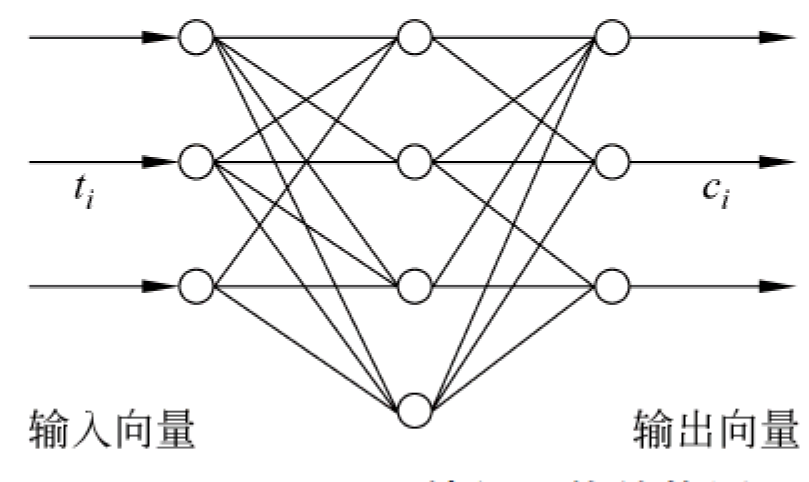


图 6-8　三层 BP 神经网络结构圈

对于三层 BP 神经网络，其输入向量为 $\boldsymbol{d}=(t_1,t_2,\cdots,t_n)$，输出向量为 $\boldsymbol{C}=(c_1,c_2,\cdots,c_m)$，输入层为 n 个神经元，隐藏层为 h 个神经元，输出层为 m 个神经元。n 为输入向量维数，m 为输出向量维数，隐藏层的神经元个数 h 可认为与问题相关，目前的研究结果还难以给出 h 与问题的类型和规模之间的函数关系。输入层和隐藏层之间、隐藏层和输出层之间的连接权重在神经网络的训练阶段，根据训练样本学习得到。给定一段文本及其特

征集，输入层神经元的个数设定为特征集的大小，输出层神经元的个数设定为类别集的大小，定义该神经网络的输入向量第 i 个分量的值为：当文本中存在特征集中的第 i 个特征词时 $t_i=1$，反之为 0。

在训练神经网络的时候，定义输出向量第 j 个输出值：当文本属于类别集中的第 j 类时 $c_i=1$，反之为 0。

使用 BP 算法进行训练，当网络稳定下来后，节点间的权值就作为文本分类时的知识，利用它完成文本分类的任务。

3. 基于规则的方法

1）决策树方法

决策树是一种多级分类方法，利用树把一个复杂的多类别分类问题转化为若干个简单的分类问题来解决。它采用分级的形式，使分类问题逐步得到解决。另外，决策树很容易转化成分类规则。一般来说，一个决策树由一个根节点 n、一组非终止节点 n_i 和一些终止节点 t_i 组成，可对 t_j 标以各种类别标签。有时不同的终止节点上可以出现相同的类别标签。如果用 T 表示决策树，那么，一个决策树 T 对应于特征空间的一种划分，它把特征空间划分成若干个区域，在每个区域中，某个类别的样本占优势，因此，可以标以该类样本的类别标签。

目前已经进行了大量的关于决策树分类研究工作，这些研究工作涉及决策树推导、决策树属性选择、决策树裁剪、由决策树抽取分类规则、提高决策效率、提高决策树的扩展性等，并开发了很多基于决策树的分类算法和系统等。

2）基于关联规则的分类法

迄今为止，基于关联规则分类方法的研究还不是太多，为人们所知晓的有三个典型的基于关联规则的分类方法：关联规则聚类系统（Association Rule Clustering System，ARCS）、关联性分类和聚合模式分类（Classification by Aggregating Patterns，CAEP）。

ARCS 方法首先使用聚类技术获得关联规则，然后使用这些规则进行分类。

关联性分类中挖掘的规则的形式为：

$$\text{condest}?y$$

其中，condset 是一组属性名和值对的集合，y 是类别标签。满足最小支持度的规则被认为是频繁的，满足最小可信任的规则被认为是准确的。假如一组规则含有相同的 condset，那么具有最高可信度的规则被选出作为可能规则（Possible Rule，PR），并代表这组规则。关联性分类方法包括两个阶段：第一阶段，寻找所有既频繁、又准确的 PR 的集合；第二阶段，利用已发现的 PR，采用一个启发式方法来组建。

CAEP 使用子集支持度的概念，来挖掘新兴模式，并用新兴模式来组建分类器。

3）粗糙集

粗糙集理论主要用于分类过程中发现非准确数据或噪声数据中的结构关系。仅适用于离散属性数据，因而，连续属性的数据必须先进行离散化处理，才可以基于粗糙集理论进行分类。

粗糙集理论基于在给定的训练数据中构建等价类。就描述数据的属性而言，一个等

价类中的所有数据样本是一致的、等同的。粗糙集能够用于近似地或“粗略”地定义这些类。对于给定的类 C 的粗糙集为两个近似集合 C 的下近似和上近似(lower and upper approximation)。C 的下近似中的数据样本毫无疑问、绝对属于类别 C,而 C 的上近似中的数据样本不能被说明为不属于类别 C。可以为每个类生成决策规则,典型的情况是,用决策表来表示规则。

粗糙集还用于降维处理和相关性分析。寻找能够描述给定数据集中所有概念的属性的最小集合是个 NP 难问题,然而,已经提出了用于降低计算强度的算法,例如,一种算法中采用可辨别矩阵(discernibility matrix)不是搜索整个训练集,而是在矩阵中搜寻冗余属性。

6.5 文本隐写分析技术

6.5.1 文本隐写分析技术概述

隐写分析技术(steganalysis),或称为隐藏信息分析技术,是对隐写术的分析和攻击技术的研究。隐写分析技术的提高有利于防止隐写术的非法应用,可以起到防止机密资料流失、揭示非法信息、打击恐怖主义、预防灾难发生的作用,从而保证国家的安全和社会的稳定。对隐写分析的研究不仅具有重要的应用价值,还具有重要的学术意义。隐写分析研究可以揭示当前信息隐藏技术的缺陷,对信息隐藏算法的安全性进行测试与评价,这是信息隐藏技术发展与完善的一条有效途径。

文本由于具有编码简单、使用灵活等特点,已成为互联网中最常见的一种信息载体。大多数的杂志、报纸、科学刊物和会议都提供了数字文档,随着电子商务及电子政务的快速发展,党政机关、企事业单位、民间团体、国防、国家安全等部门将有大量的文字材料通过互联网传输。由于文本的易传播和易编辑性,一些不法分子利用文本作载体进行隐蔽通信。因此文本隐写分析技术作为对抗文本信息隐藏技术的主要手段日益得到重视与发展,其中尤以文本隐藏信息检测技术更为重要。

文本隐写分析技术是针对文本隐写技术进行检测的技术。目前文本隐写分析方式可分为如下几类。

1. 针对 Mimic 类型文本隐藏方法的检测

Mimic 模式由 Peter Wayner 提出,它通过使用一种被称之为 Mimic Function 的处理方法,将要隐藏的秘密信息 A 进行伪装得到 A',使得 A' 与无辜信息 B 具有相同的统计特性,使监控方的自动检测系统把 A' 误判为 B,从而达到逃避检测、保障个人私密通信安全的目的。由于该模式具有实现比较容易、抗检测能力强等优点,现已成为基于自然语言处理的文本隐写方法中一种很常用的模式。

1) 利用检测熵,采用支持向量机分类的检测方法

利用检测熵,采用支持向量机分类的检测方法首先定义单词 x 的 score 值:$S_x =$

$\frac{1}{c}\left(\sum_{i=1}^{n} i\right)$，其中，$n$ 表示文中单词的总数。假设一篇文章中有 N 个不同的词，定义检测熵为：$DE=\sum_{i=0}^{N-1} S_i \log \frac{1}{S_i}$，同时检测熵的方差为：

$$\mathrm{VAR(DE)}=\sum_{i=0}^{N-1} S_i\left(\log \frac{1}{S_i}-\mathrm{DE}\right)^2$$

将 DE 和 VAR(DE) 作为特征，采用支持向量机分类。这种方法可以检测采用了 Nicetext 和 Texto 等隐写工具隐藏的信息。

2）利用分布度，基于支持向量机的检测方法

利用分布度，基于支持向量机的检测方法提出，对于自然文本，某个重复的单词通常有着不平衡的分布，也就是说，这些重复单词出现的地方一般比较集中，而对于隐写文本来说，这些位置具有随机性，相对分散。假设一篇文章中的单词序列为 $S=\{w_0,w_1,\cdots,w_n\}$，单词 w_i 的位置定义为：$wl_i=\frac{i}{n}$，假设单词 w_k' 出现的次数为 w_k，定义单词 w_k' 的分布度：

$$\mathrm{SD}(w_k')=\frac{1}{n_k}\left(\sum_{i=0}^{n_k} wl_{ki}-\mathrm{Avg}(w_k')\right)^2$$

其中，$\mathrm{Avg}(w'_k)=\frac{1}{n_k}\sum_{i=0}^{n_k} wl_{ki}$。计算一篇文章分布度的均值和方差：

$$\overline{\mathrm{SD}}=\sum_{i=0}^{m} \mathrm{SD}(w_i')\frac{n_i}{n}$$

$$\mathrm{VAR(SD)}=\sum_{i=0}^{m} \mathrm{SD}(w_i'-\overline{\mathrm{SD}})\frac{n_i}{n}$$

作为支持向量机的输入特征，一次区分正常文本和隐写文本。

3）基于建立语言统计模型的检测方法

计算 n 个词连续出现的最大似然概率：$P(w_n|w_1\cdots w_n)=\frac{C(w_1\cdots w_n)}{C(w_1\cdots w_{n-1})}$，然后计算一篇文章的 P 值：

$$P(w)=\sqrt[N]{\prod_{i=1}^{N}\frac{1}{P(w_i \mid w_{i-2}w_{i-1})}}$$

根据 P 值的不同区分正常文本和隐写文本。

4）基于句间相关性度量判断法

基于句间相关性度量判断法对现在比较流行的语义隐藏算法弱点进行分析，总结出两种导致语义隐藏技术泄露的情况：掩体文本选择不当和嵌入文本信息容量较大。如果选择的文本载体中含有一些术语或限定性很强的词，由于这些词与所在的短语或句子紧密相连，稍微改动语序或替换词语都将使得原有句子含义发生变化甚至不可理解。嵌入文本越多导致产生的隐写文本也就越多，进而导致一个句子的掩体文本已经无法容纳嵌入文本。由于掩体文本的句子是独立产生的因此句子间不相关或相关性很弱，从而导致

隐藏信息的暴露。根据这些分析，作者首先通过概念图中的最大连接匹配得出的最大连接数的值与相容节点值累加得到句间相关性度量，然后根据设定的门限判断是否有隐藏信息存在。

5）基于首字母分布的检测方法

基于首字母分布的检测方法由公式

$$\alpha = 1 - \frac{1}{2}\sum_{i=1}^{26} | f_1(i) - f_2(i) |$$

计算出 α 并通过 α 来判断是否是隐写文本，其中，$f_1(i)$ 和 $f_2(i)$ 分别为待检测文本与字典中首字母序号为 i 的所有单词的出现频率。该方法认为：因为嵌入的秘密信息是伪随机的，那么如果使用某个字典 D 隐藏信息后得到了隐写文本 T，则 T 中单词出现的频率必然与 D 中的单词频率接近，T 中单词按首字母序号分类后每类的频率应该也与 D 中的单词对应类的频率近似，即应该有 $f_T(i) \approx f_D(i), i \in [1,26]$，此时求得的 α 值应该比较大；而自然语言文本中的单词频率则会与字典 D 中对应的单词频率相差较远，此时求得的 α 值应该相对较小；所以如果 α 大于某个阈值则可以认为是隐写文本。

6）基于文本剩余度的检测方法

基于文本剩余度的检测方法把文本作为 m 阶时齐马尔可夫信源，文本中的单词作为信源符号，用 m 阶时齐马尔可夫信源的特性计算得到这个文本中连续 m 个单词之间的一些相关特性，然后根据 mimic 模式产生的隐写文本的文本剩余度比正常文本高这一规律对文本进行检测。

2. 针对基于同义词替换的文本隐藏方法的检测

1）基于支持向量机的文本隐藏信息检测方法

基于支持向量机的文本隐藏信息检测方法，首先通过训练正常文本和使用基于同义词替换的文本隐写方法嵌入秘密信息后的隐写文本的语言模型来获取正常文本和隐写文本的模式；然后，基于从语言模型获得的统计输出训练一个支持向量机分类器；最后，对于一个给定的文本，可以基于这个 SVM 分类器的输出来确定是否是使用基于同义词替换的文本隐写方法嵌入了隐藏信息的文本。

2）利用 Internet 统计上下文搭配的方法

利用 Internet 统计上下文搭配的方法首先定义包含单词 w 的文本数量 $\mathrm{CF}(w)$ 以及包含集合 S 中的所有的单词的文本数 $\mathrm{CF}(S)$；然后定义适合度：$\mathrm{ST}(w,C)=\ln \dfrac{N}{\mathrm{CF}(w)} \cdot \mathrm{CF}(w,C)$；最后将适合度的均值和方差作为支持向量机的特征，对正常文本和隐写文本分类。

3）基于分析文本中同义词结对值来进行这类隐藏信息的检测

首先把任一文本 D 中所有表示语义 y 的词按在 D 中的出现顺序排列，得到一个长为 l 的序列 $w_1, w_2, \cdots, w_l$，如果 $l \geqslant 3$，则另任取 $w_0 \neq w_1, w_l \neq w_{l+1}$ 构成一个新序列 $w_0, w_1, \cdots, w_{l+1}$，如果在这个新序列中有 $w_{i-1} \neq w_i = w_{i+1} \neq w_{i+2}$（其中 $i \in [1, l-1]$），则称文本 D 中在一个 S_y 下的同义词结对。文本 D 中的同义词结对总数称为 D 中的同义词结

对值。然后分析得出结论：隐写文本的结对值要大于正常文本。因此，根据一篇文本的结对值，我们可以判断它是隐写文本还是正常文本。

当前文本隐藏信息检测技术往往存在以下缺点：

(1) 难以检测基于自然语言处理技术的文本隐写方法。由于基于自然语言处理技术的文本隐写方法的特殊性，对基于自然语言处理技术的文本隐写方法的检测比较困难。现有的检测方法一般是依靠自然语言处理技术，使用庞大的语料库来分析被检测文本与自然语言文本的近似程度，通常不仅实现难度比较大，而且由于目前自然语言处理技术并不完善，使得这些检测方法受自然语言处理技术的干扰比较大，往往难以取得较好的结果。

(2) 难以发现大文本载体中的小隐藏信息。这是所有隐藏信息检测中都存在的现象，大载体中的小隐藏信息容易被载体本身的干扰信息掩盖，所以通常难以检测。

(3) 有些检测算法缺乏实用性。例如美国普渡大学信息安全教育与研究中心提出的基于支持向量机的文本隐藏信息检测方法，虽然进行了实现，但是该方法实现难度高，而且能够提供的数据有限，所以其他人难以重复该实验；还有一些检测算法由于成功率偏低而缺乏实用价值。

(4) 检测方法的通用性不强。当前文本隐藏信息检测技术通常是针对某一种具体的文本隐写方法而言的，甚至是针对某一个具体的文本隐写工具软件而言的，能检测的文本隐写方法比较单一，缺乏通用性。但是随着文本隐写方法的增多，如何能用尽可能少的检测方法检测到尽可能多的隐写方法就变得越来越重要，因为如果对于每种不同的隐写方法都使用不同的检测算法的话，多种检测方法集成在一起使用时总体虚警率有可能会存在累加情况，从而导致检测系统的总体虚警率居高不下。

(5) 往往只考虑了文本隐写分析中的检测，而没有考虑到文本隐写分析中的提取与恢复。

6.5.2 典型的文本隐写分析方法

基于同义词替换的文本隐写方法是文本隐写中的一种典型方法，它可以通过对载体中的同义词进行有选择的替换来嵌入隐藏信息，该方法嵌入隐藏信息后会导致载体文本中同义词结对概率的明显增加。基于此，本节介绍一种通过分析文本中同义词结对值来进行隐藏信息检测的方法。

对于普通的自然语言文本，其中的词通常可以用对应的同义词替换掉，这种替换既不会引起该文意义的明显改变，也不会对该文的句法结构造成任何影响。基于同义词替换的文本隐写方法就是利用这个特性，在文本中通过对选择的同义词进行替换来嵌入隐藏信息。

对于任一文本 D，把 D 中所有表示语义 y 的词按在 D 中的出现顺序排列，得到一个长为 l 的序列 $w_1, w_2, \cdots, w_l$，如果 $l \geqslant 3$，则另任取 $w_0 \neq w_{i+2}$（其中 $i \in [1, l-1]$），则称文本 D 中在 w_i 处存在一个 S_y 下的同义词结对。

设 WL 为文本所属语种包含的所有词；$M(x)$ 为 x 的词义，其中 $x \in W$。设有集合满

足 $S_y=\{x|M(x)\approx y,x\in ML\}$，则称 S_y 为一个词义为 y 的同义词组，该组中包含了 $|S_y|$ 个具有相同或相近的词；即同义词组为具有相同或相近意义的一组词的集合。选定 z 个不同的同义词组组成集合 DB，即 $DB=\{S_{y1},S_{y2},\cdots,S_{yz}\}$，对每个同义词组中的词进行编码。这里 DB 就称为一个同义词库。

1. 嵌入过程

依据嵌入算法找到需要嵌入的数据对应位置的词 w，设 $M(w)=k$，则 w 所属的同义词组为 $S_k(S_k\in DB)$，设 w 在 S_k 中对应的编码为 c，需要嵌入在该位置的数据为 d；如果 $c\neq d$，则在 S_k 中找到编码为 d 的词 w'，并用 w' 在载体中当前位置替换掉 w，如果 $c=d$ 则不进行替换；继续查找和替换下一个需要嵌入的位置；直到嵌入结束。

2. 提取过程

依据提取算法找到需要提取的数据对应位置的词 w，设 $M(w)=k$；把 w 在 S_k 中的编码 c 作为该位置的提取数据；继续查找和提取下一个需要提取的位置；直到提取结束。当前已有的基于同义词替换的文本隐写算法的嵌入与提取基本上都可以用上面这个嵌入与提取过程描述。

3. 基于同义词替换的文本信息隐藏方法的检测算法

首先提出的是文本中同义词结对值的获取算法，然后在此基础上将提出一种基于同义词替换的文本信息隐藏方法的检测算法。

1）结对值的获取算法

输入：待检测的文本 D、同义词库 DB。

输出：文本 D 中的同义词结对值。

① 使用分词系统得到 D 经过分词后的词的序列 $W=W_0,W_1,\cdots,W_{num-1}$，其中，num 为 D 分词后词的总数，$i=0$，$n=0$。

② 如果 $i\geqslant num$，则跳到步骤④；获取词 W_i 在 DB 中对应的组名：TName = FindInDb(W_i)；如果 DB 中存在 W_i，则进入步骤③，否则 $i=i+1$，重复本步骤。

③ 临时记录表 TT 中包含的序列数为 n，计算 k = FindIndexInTT(TName)；如果 $k\in[0,n-1]$，则 $Team_k$ 为临时记录表 TT 中 Tname 对应的序列；如果 $k<0$，则取 $k=n$，并在 TT 中新建一个序列 $Team_k$ 与这个 TName 对应，$n=n+1$。

将 W_i 添加到 $Team_k$ 尾部；$i=i+1$，转到步骤②。

④ 依次对 TT 中的每个 $Team_k$ 进行结对值统计分别得到 g_k，其中 $k\in[0,n-1]$。

⑤ 返回待检测的文本 D 的结对值为 $NSP_D=\sum_{k=0}^{n-1}g_k$。

由离散型随机变量的分布函数：

$$F(x)=p(\xi<x)=\sum_{x_i<x}p(\xi=x_i)$$

可得到文本中同义词结对值分布曲线。

在任一文本 D 中，如果出现了某组同义词 S_y 下的结对，那么该组同义词称为 D 中的结对相关组。文本 D 中结对相关组的数量称为 D 中的结对相关组数。

2）基于同义词替换的文本信息隐藏方法的检测算法

输入：待检测的文本 D。

输出：是否包含隐藏信息。

由结对值的获取算法可知文本 D 中出现同义词的总组数 $N_D=n$；而文本 D 中结对相关组数为：$C_D=\sum_{k=0}^{n-1}\mathrm{sgn}\,(g_k)$。如果$(C_D/N_D)\geqslant\sigma$，则返回结果为True，否则返回结果为False。

思　考　题

6.1　简述自然文本的分布特征及常见的自然语言处理技术。

6.2　简述文本数字水印常见算法。

6.3　简述文本隐写术与水印技术的异同。

6.4　什么是文本过滤和文本分类？二者之间有什么联系？

6.5　谈谈你对文本隐写分析的理解。

6.6　如何实现对文本内容的加密？

参考文献

[1]　方滨兴.关于信息安全属性可计算性初探.中国信息安全大会，2006

[2]　曹卫兵，戴冠中，夏煜，等.基于文本的信息隐藏技术.计算机应用研究，2003，20(10)39-41.

[3]　Kazenbeisser S，Petitcolas F A P. Information Hiding Techniques for teganography and Digital Watermarking. Canton St. Norwood，A USA：Artech House Publisers，2000.

[4]　黄华，齐春，李俊等.一种新的文本数字水印标记策略和测方法.西安交通大学学报，2002，36(2) 165-168.

[5]　Atallah M J，Raskin V，Christian F，et al. Natural Language Watermarking and Tamperproofing. Proceedings of Information Hiding 5th International Workshop，2002：196-212.

[6]　Atallah M J，Raskin V，Crogan M，et al. Natural Language Watermarking：Design，Analysis，and a Proof of Concept Implementation. Proceedings of Information Hiding 4th International Workshop，Pittsburgh，PA，USA，Berlin：Springer，2001-04：185-199.

[7]　白剑，杨榆，徐迎晖，等.基于文本的信息隐藏算法.计算机系统应用，2005，(4)，32-35.

[8]　王智，周洪玉.基于 Word 文档的信息隐藏方法的实现.信息技术，2008，(11)：30-31.

[9]　耿红琴.基于文本的信息隐藏技术研究.科学技术与工程，2007，7(6)：1070-1073.

[10]　李丽娟，熊淑华.基于文本的信息隐藏技术研究.现代电子技术，2006，(5)：67-69.

[11]　Brassil J，Low S，Maxemchuk N，et al. Document marking and identification using both line and word shifting. Technical Report，AT&T Bell Laboratories，1994.

[12]　罗纲，孙星明，刘玉玲.基于噪声检测的文本隐藏信息检测算法研究.湖南大学学报(自然科学

版),2005,32(6.增刊):181-184.

[13] Katzenbeisser S. Principles of steganography. In: Proc of Information Hiding: Techniques for Steganography and Digital Watermarking. Boston: Artech House,2000,17-41.

[14] Topkara M,Topkara U,Atallah M J. Information hiding through errors: a confusing approach. In: Proceedings. of SPIE Conference on Security, Steganography, and Watermarking of Multimedia Contents Ⅸ. San Jose,2007,65050V.

[15] Bender W,Gruhl D,Morimoto N,et al. Techniques for data hiding. IBM Systems Journal,1996, 35(3&4): 313-336.

[16] Atallah M J, McDonough C J, Raskin V, et al. Natural language processing for information assurance and security: An overview and implementations. In: Proceedings of the 2000 workshop on new security paradigms. New York: ACM Press,2000,51-65.

[17] 吴悠,孙星明.基于正弦波的 Word 文档数字水印.计算机工程,2005,31(24):175-176,209.

[18] 张静,张春田.用于 PDF 文档认证的数字水印算法.天津大学学报,2003,36(2),215-219.

[19] 刘友继,孙星明,罗纲.一种新的基于 PDF 文档结构的信息隐藏算法.计算机工程,2006,32(17): 230-232.

[20] 廖继旺,孙洪淋.中文 Word 文档中数字水印的设计与实现.科学技术与工程,2006,6(7), 877-879.

[21] Zhao Q J,Lu H T. A PCA-based watermarking scheme for tamper-proof of web pages. Pattern Recognition,2005,38(8),1321-1323.

[22] Yao R H,Zhao Q J,H. T. Lu. A novel watermark algorithm for integrity protection of XML documents. International Journal of Computer Science and Network Security, 2006, 6 (2), 202-207.

[23] Zhou X, Pang H H, Tan K L, et al. WmXML: A system for watermarking XML data. In: Proceedings of the 31st VLDB Conference. Norway,2005,1318-1321.

[24] Huang D,Yan H. Interword distance changes represented by since waves for watermarking text images. IEEE Trans on Circuits and System for Video Technoloy,2001,11(12): 1237-124.

[25] Borges P V K,Mayer J. Document watemarking via character luminance modulation. Proceedings of IEEE International Conference on Aeoustics, Speech and Signal Processing, ICASSP2006. Toulouse,Franee. 2006.

[26] Shirali-Shahreza M H, Shirali-Shahreza M. A new approach to Persian/Arabic text steganography: computer and information science. Proceedings of the 5th IEEE/ACIS Imitational Conference,Honohilu,HI,US. 2006.

[27] Thiemert S,Steinebaeh M,P. Wolf. A digital watermark for vector-based fonts. Proceedings of the 8th ACM Multimedia and Security Workshop,Geneva,Switzerland,2006.

[28] 罗纲,孙星明,向凌云等. 针对同义词替换信息隐藏的检测方法研究. 计算机研究与发展,2008, 45(10):1696-1703.

[29] 罗纲,孙星明.基于文本剩余度的文本隐去信息检测方法研究.通信学报,2009,30(6): 19-25.

[30] 赵敏之,孙星明,向华政.基于虚词变换的自然语言信息隐去算法研究.计算机工程与应用,2006, 42(3): 158-160.

[31] Sui X G, Luo H. A Steganalysis method based on the distribution of space characters. Proceedings of 2006 International Conference on Communications, Circuits and Systems, Guilin, China, 2006, 54-56.

[32] Chen Z L，Huang L S.，Yu/Z S，et al. Linguistic stegangraphy detection using statistical characteristics. Proceedings of Information Hiding 2008，LNCS 5284，2008：224-235.

[33] Xiong L Y，Sun X M，Luo G，et al. Research on Steganalysis for text steganography based on font format. Proceedings of the Third International Symposium on Information Assurance and Security(IAS'07)，IEEE Press，Manchester，UK，2007：490-495.

[34] Liu Y L，Sun X. M，Ingemar J Cox，et al. Natural Language information hiding based on Chinese Mathematical Expression. Interational Journal of Network Seurity，2009，8(1)：10-15.

第 7 章 chapter 7

数字图像内容安全

本章学习目标

数字图像是最为常见的数字内容之一，本章将针对数字图像的特点，介绍数字图像以及数字图像内容的相关概念，对数字图像加密技术、数字图像水印技术以及数字图像隐写分析技术进行深入的阐述。

通过本章的学习，应掌握以下内容：

(1) 数字图像及数字图像内容安全的相关概念。

(2) 数字图像编码方式。

(3) 数字图像加密技术。

(4) 数字图像水印技术。

(5) 数字图像隐写分析技术。

7.1 数字图像内容安全基本概念

7.1.1 数字图像的概念、分类及特点

1. 图像、图形和数字图像

1) 图像与图形

图像是当前最为常见的信息表达方法之一，它是对客观世界的反映。“图”是指物体透射或反射光的分布。图像既是一种光的分布，也包含人的视觉心理因素。图像的最初取得是通过对物体和背景的“摄取”。这里的“摄取”即意味着一种“记录”过程，如照相、摄像、扫描等，这是图像和图形的主要区别。人们对图像是很熟悉的，生活中人们很容易说出哪些东西是图像，图像是人对视觉感知的物质的再现。图像可以由光学设备获取，如照相机、镜子、望远镜、显微镜等；也可以人为创作，如手工绘画、计算软件生成图像等。

图形是用数学规则产生的或具有一定规则的图案。图形往往是用一组符号或线条来表示的。例如房屋设计图，我们是用线条来表示房屋的结构。

2) 数字图像

任意一幅数字图像粗看起来似乎是连续的，实际上是不连续的，它由许多密集的细

小点所组成，这些细点构成一幅图像的基本单元，称为像素。就像任何物质一样，肉眼看上去是连续的，但实质上都是由一个一个分子组成。显然点越多，像素越多，画面就越清晰。日常所见的图像许多是连续的，为了能用计算机对图像进行加工，需要把连续的图像在坐标空间 XY 和性质空间 F 都离散化。这种离散化的图像就是数字图像，它是客观事物的可视数字化的表达。数字图像可用 $I(r,c)$来表示，其中，I、r、c 的值都是整数。这里 I 代表离散化后的 F，(r,c)代表离散化后的(x,y)，其中，r 代表图像的行，c 代表图像的列。

2. 数字图像的表示方式

1）图像的矩阵和矢量表示

一幅 2D 图像可以用一个 2D 数组 $f(x,y)$来表示。实际中还常将一幅 2D 图像写成一个 2D 的 $M\times N$ 矩阵(其中 M 和 N 分别为图像的总行数和总列数)。

$$F=\begin{bmatrix} f_{11} & f_{12} & \cdots & f_{1N} \\ f_{21} & f_{22} & \cdots & f_{2N} \\ \vdots & \vdots & & \vdots \\ f_{M1} & f_{M2} & \cdots & f_{MN} \end{bmatrix} \tag{7-1}$$

一幅 2D 图像也可以用矢量来表示，可写成：

$$\boldsymbol{F}=[f_1 \quad f_2 \quad \cdots \quad f_N] \tag{7-2}$$

其中，

$$f_i=[f_{1i} \quad f_{2i} \quad \cdots \quad f_{Mi} \quad]^{\mathrm{T}}, \quad i=1,2,\cdots,N \tag{7-3}$$

上述两种表示形式可以方便地互相转换。对于 $M\times N$ 像素的彩色图像，可以用三个矩阵表示：$[F_R]_{M\times N}$、$[F_G]_{M\times N}$、$[F_B]_{M\times N}$，其中 R、G、B 代表彩色图像的三个基本的颜色通道。

2）数字图像的种类和显示方式

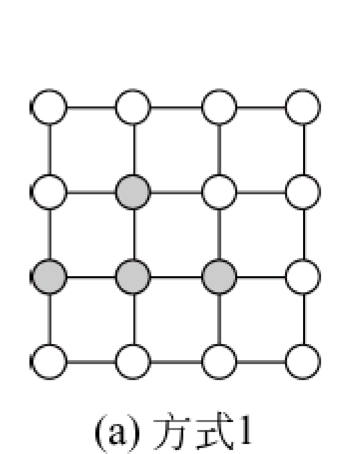
每个图像的像素通常都对应于二维空间中一个特定的位置，并且由一个或者多个与那个点相关的采样值组成数值。根据这些采样数目及特性的不同数字图像可以划分为二值图像、灰度图像和彩色图像。二值图像(binary image)中每个像素的亮度值(intensity)仅可取 0 和 1。灰度图像(gray scale image)中每个像素可以由 0(黑)～255(白)的亮度值表示。0～255 之间表示不同的灰度级。彩色图像(color image)，即每幅彩色图像是由三幅不同颜色的灰度图像组合而成，分别是红色(Red)、绿色(Green)和蓝色(Blue)。

对 2D 图像的显示可以采用多种形式，一般是将 2D 图像看作在 2D 空间位置上的一种幅度分布。根据图像的不同，采取的显示方式也可不同。例如对二值图像，在每个空间位置的取值只有两个，可用黑白来区别，也可用 0 和 1 来区别，如图 7-1 所示。

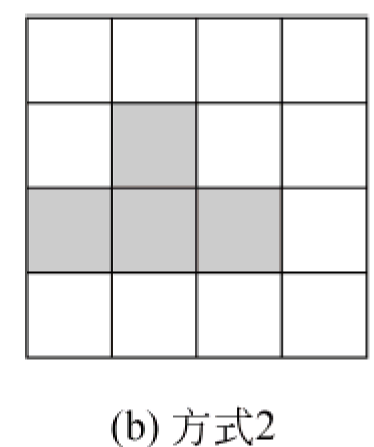
在图中，各种方式表示的都是一个 4×4 的二值图像，一般说这幅图像的空间分辨率是 4×4，也就是说这幅图像在空间有 4×4＝16 个位置可显示独立的灰度。

3）数字图像的格式

数字图像的格式是人们保存图像的一种方式或形式。目前比较流行的图像格式包



(a) 方式1



(b) 方式2

0	0	0	0
0	1	0	0
1	1	1	0
0	0	0	0

(c) 方式3

图 7-1　3 种表达同一个 4×4 的二值图像矩阵的可视方式

括光栅图像格式 BMP、GIF、JPEG、PNG 等，以及矢量图像格式 WMF、SVG 等。其中数码相机的保存格式大多是 JPEG 格式，大多数浏览器都支持 GIF、JPG 以及 PNG 图像的直接显示。而矢量图像却通常需要专门的软件才能显示，其中 SVG 格式作为 W3C 的标准格式在网络上的应用越来越广。但是，并不是所有图像格式都适合进行数字图像处理，例如，GIF 是经过压缩的图像格式，在很多情况下都不适合进行数字图像处理。进行数字图像处理的前提是图像必须是未经过压缩的，最常见的 BMP 格式是未经压缩的，所以适合作为处理对象。压缩图像必须经过解压缩还原成 BMP 图像才能用于数字图像处理。

3. 数字图像的性质与特点

像素（像元）是数字图像最基本的单位，是成像过程的采样点，也是计算机图像处理的最小单元。

通常，数字图像具有以下特点：

（1）图像数据信息量很大。例如取 512×512 个像素组成一幅数字图像，如其灰度级用 8 比特的二进制来表示，则有 $2^8=256$ 个灰度级，那么这幅图像的数据信息量即为 512×512×8＝2097152 比特。若是彩色图像，数据量会更大。对这样大数据量的图像进行处理，必须使用大内存的计算机才能胜任。如果计算机的性能达不到一定的要求，则无法有效处理数字图像。

（2）数字图像占用的频带较宽。与语言信息相比，占用的频带要大几个数量级。如电视图像的带宽为 5.6MHz，而语言带宽仅为 4kHz 左右。频带越宽，技术实现的难度就越大，成本亦越高，为此对频带压缩技术提出了较高的要求。

（3）数字图像中各个像素不是独立的，其相关性很大。也就是说，在图像中通常有大块区域的灰度值是相差不大的。例如在一幅数字电视图像中，同一行中相邻两个像素或相邻两行的像素，其相关系数可达 0.9，而相邻两帧电视图像之间的相关性比帧内相关性还要大一些，因此图像信息的冗余度很大。

（4）数字图像是需要给人观察和评价的，因此效果的好坏受人的因素影响较大。由于人的视觉系统比较复杂，数字图像受环境条件、视觉性能、人的主观意识的影响很大，因此要求系统与人必须有良好的配合，这还是一个很大的研究课题。

7.1.2　数字图像的编码方式

随着各种现代化技术的不断发展，图像信息已变为众多应用领域的重要处理对象，

怎样充分利用信道传输有用的图像信息就成了一个现实问题。如果要对原始图像进行存储、记录和传输,那么必须要对数字图像的信息进行有效的压缩。随着各种技术的不断发展,数字图像的数据压缩在数字图像传输中发挥着关键性的作用。

1. 图像编码的基本原理

虽然表示图像需要大量的数据,但是图像数据是高度相关的,或者说存在冗余信息,去掉这些信息后可以有效压缩图像,同时不会损害图像的有效信息。数字图像的冗余主要表现为以下几种形式:空间冗余、时间冗余、视觉冗余、信息熵冗余、结构冗余和知识冗余。图像数据的这些冗余信息为图像压缩编码提供了依据。图像编码的目的就是为了充分利用图像中存在的各种冗余信息,特别是空间冗余,时间冗余以及视觉冗余,以尽量少的比特数来表示图像。利用各种冗余信息,压缩编码技术能够很好地解决将模拟信号转换为数字信号后所产生的带宽需求增加的问题,它是推动数字信号走上实用化的关键技术之一。

图像编码主要是利用图像信号的统计特性以及人类视觉的生理学及心理学特性,对图像信号进行高效编码,即研究数据压缩技术的目的是在保证图像质量的前提下压缩数据,便于存储和传输,以解决数据量大的矛盾。一般来说,图像编码的目的有三个:

(1) 减少数据存储量。

(2) 降低数据率以减少传输带宽。

(3) 压缩信息量,以便于特征提取,为后续识别做准备。

2. 经典的编码技术

经典图像编码技术根据编码原理可分为熵编码、预测编码、变换编码和混合编码等。

1) 熵编码

熵编码纯粹基于信号统计特性的编码技术,是一种无损编码。熵编码的基本原理是给出现概率较大的符号赋予一个短码子,而给出现概率较小的符号赋予一个长码字,从而使得最终的平均码长很小。常见的熵编码方法有行程编码、霍夫曼编码和算术编码等。

(1) 行程编码

行程编码又称行程长度编码,是一种熵编码,该编码属于无损压缩编码。对于二值图像有效。其基本原理是:将具有相同值的连续串用其串长和一个代表值来代替,使符号长度少于原始数据的长度。改变连续串就称为行程,串长称为行程长度。

例如,666666688888822333355555 的行程编码为(6,7)(8,6)(2,3)(3,4)(5,5)。可见,行程编码的位数远远少于原始字符串的位数。

行程编码分为定长和不定长编码两种。定长编码是指编码的行程长度所用的二进制位数固定,而变长行程编码是指对不同范围的行程长度使用不同位数的二进制位数进行编码。使用变长行程编码需要增加标志位来表明所使用的二进制位数。行程编码比较适合于二值图像的编码,一般用于量化后出现大量零系数连续的场合,用行程来表示连零码。行程编码对传输差错很敏感,一位符号出错就会改变行程编码的长度,使整个

图像出现偏移，因此，一般要用行同步、列同步的方法，把差错控制在一行一列之内。它适用于那些包含很少灰度级的图像，对单一颜色背景下物体的图形图像可以达到很高的压缩比，但对其他类型的图像压缩比就很低。

(2) 霍夫曼编码

霍夫曼(Huffman)编码是1952年为文本文件而开发出的一种熵编码，属于无损压缩编码。该方法完全依据字符出现的概率来构造码字，对频繁出现的字符多使用较短的码字，而对出现次数较少的字符使用较长的码字。在具有相同信源概率分布的前提下，它的平均码字长度是最短的。变长最佳编码定理是霍夫曼编码的理论基础。

静态霍夫曼编码使用一棵在压缩之前就建好的编码树，它是根据字符出现的概率来生成的。相反，动态霍夫曼编码是在编码过程中建立它的编码树。具体的方法是，在分配码字长度时，首先将其中概率最小的两个符号的概率求和，并把它看作是一个新组合符号的概率，再与其他符号按概率递降顺序排列，重复上述做法，直到最后只剩下两个符号的概率为止。然后开始以相反顺序逐步进行编码，每一步有两个概率分支，各赋予一个二进制的码。可以对概率小的赋编码为0，概率大赋1，也可以反过来赋编码。这种统计方法能够达到更高的压缩比，而且此方法简单有效，编码效率高。但是，这是以增大编码和解码的时间为代价的。

霍夫曼编码具有一些明显的特点：

① 编出来的码都是异字头码，保证了码的唯一可译性。

② 由于编码长度可变，因此译码时间较长，使得霍夫曼编码的压缩与还原相当费时。

③ 编码长度不统一，硬件实现有难度。

④ 对不同信号源的编码效率不同，当信号源的符号概率为2的负幂次方时，达到100%的编码效率；若信号源符号的概率相等，则编码效率最低。

⑤ 由于0与1的指定是任意的，故由上述过程编出的最佳码不是唯一的，但其平均码长是一样的，故不影响编码效率与数据压缩性能。

(3) 算术编码

算术编码是20世纪80年代发展起来的一种熵编码方法，其基本原理是将被编码的数据序列表示成0和1之间的一个间隔(也就是一个小数范围)，该间隔的位置与输入数据的概率分布有关。信息越长，表示间隔就越小。因而表示这一间隔所需的二进制位数就越多。算术编码有两种模式：一种是基于信源概率统计特性的固定编码模式，另一种是针对未知信源概率模型的自适应模式。

算术编码适合于由相同重复序列组成的文件，算术编码接近压缩的理论极限。这种方法将不同的序列映像到0到1之间的区域内，该区域表示成可变精度(位数)的二进制小数，越不常见的数据需要的精度越高(更多的位数)，这种方法比较复杂，因而并不常用。

2) 预测编码

预测编码是基于图像数据的空间或时间冗余特性，用已传输的像素对当前的像素进行预测，然后对预测误差进行量化和编码。如果预测比较准确，误差就会很小。在同等精度要求的条件下，就可以用比较少的比特进行编码，达到压缩数据的目的。预测编码

可以分为一维预测(行内预测)、二维预测(帧内预测)和三维预测(帧间预测)。常用的预测编码有差分脉冲编码调制(DPCM)和自适应差分脉冲编码调制(ADPCM)等。

(1) 差分脉冲编码调制

在PCM系统中,原始的模拟信号经过采样后得到的每一个样值都被量化成为数字信号。为了压缩数据,可以不对每一样值都进行量化,而是预测下一样值,并量化实际值与预测值之间的差值,这就是DPCM(差分脉冲编码调制)。1952年贝尔(Bell)实验室的C. C. Cutler取得了差分脉冲编码调制系统的专利,奠定了真正实用的预测编码系统的基础。在图像信号中应用DPCM时,用作预测的像素和被预测的像素可以在同一行,也可以在不同行(同一帧),甚至在不同帧,分别称为一维预测、二维预测和三维预测。

DPCM的优点是算法简单,容易用硬件实现,缺点是对信道噪声很敏感,会产生误差扩散。即某一位码出错,对图像一维预测来说,将使该像素以后的同一行各个像素都产生误差;而对二维预测,该码引起的误差还将扩散到以下的各行。这样,将使图像质量大大下降。同时,DPCM的压缩率也比较低。随着变换编码的广泛应用,DPCM的作用已很有限。

(2) 自适应差分脉冲编码调制

进一步改善量化性能或压缩数据率的方法是采用自适应量化或自适应预测,即自适应脉冲编码调制(ADPCM)。它的核心想法是:利用自适应的思想改变量化阶的大小,即使用小的量化阶去编码小的差值,使用大的量化阶去编码大的差值;使用过去的样本值估算下一个输入样本的预测值,使实际样本值和预测值之间的差值总是最小。

① 自适应量化。在一定量化级数下减少量化误差或在同样的误差条件下压缩数据,根据信号分布不均匀的特点,希望系统具有随输入信号的变化区间足以保持输入量化器的信号基本均匀的能力,这种能力叫自适应量化。

自适应量化必须具有对输入信号的幅值进行估值的能力,有了估值才能确定相应的改变量。若估值在信号的输入端进行,称前馈自适应;若在量化输出端进行,称反馈自适应。信号的估值必须简单、占用时间短,才能达到实时处理的目的。

② 自适应预测。预测参数的最佳化依赖信源的特征,要得到最佳预测参数显然是一件繁琐的工作。而采用固定的预测参数往往又得不到较好的性能。为了能使性能较佳,又不至于有太大的工作量,可以采用自适应预测。

为了减少计算工作量,预测参数仍采用固定的,但此时有多组预测参数可供选择,这些预测参数根据常见的信源特征求得。编码时具体采用哪组预测参数需根据特征来自适应地确定。为了自适应地选择最佳参数,通常将信源数据分区间编码,编码时自动地选择一组预测参数,使实际值与预测值的均方误差最小。随着编码区间的不同,预测参数自适应地变化,以达到准最佳预测。

3) 变换编码

变换编码是将空间域里描述的图像,通过某种变换(常用的是二维正交变换,如离散余弦变换、K-L变换等),映射到另一变换域中,使变换后的系数之间的相关性降低。图像变换本身并不能压缩数据,但变换后图像的大部分能量只集中到少数几个变换系数上,采用适当的量化和熵编码可以有效地压缩图像。

(1) K-L 变换

K-L 变换是一种最佳正交变换。它是用数据本身的相关矩阵对角化后完成的,这种变换将产生完全不相关的变换系数。如果图像数据之间是高度相关的,经过 K-L 变换后的系数将出现多个零值,同时某些系数的值会很小。

在 K-L 变换中不同的图像数据有不同的变换矩阵,由此造成反变换矩阵的不唯一性;另外 K-L 变换矩阵的构造计算量很大,因而它不是一种实用的变换方法,通常作为评价其他线性变换的比较基准。

(2) 离散余弦变换(DCT)

由于 K-L 变换算法复杂度较高,所以在实际编码工作中,人们常用离散余弦变换。对大多数图像信源来说,DCT 变换是现行编码方法中最接近 K-L 变换的方法。

DCT 先根据变换系数的能量分布,将整个图像分成 $N\times N$ 像素块,然后对这 $N\times N$ 像素块逐一进行 DCT 变换。其中 DCT 变换后,幅值较大的图像系数大多集中在图像块的左上角。与其他系数相比,这些低频系数包括了图像的大部分内容,所包含的能量最大,在变换图像中的地位也最重要,应使它们的量化误差最小。另一方面,大多数图像的高频分量较小,对图像质量影响甚微,加上人眼对高频成分的失真不太敏感,可以使用更粗的量化,一般采用设定阈值的方法,置小于阈值的变换系数为零,由此传送变换系数所用的码率要远远小于传送图像像素所用的码率,从而大大提高了编码效率。经区域编码和阈值编码后,变换图像的系数大部分为零,可以采用有效的方法将非零系数和零系数组织起来,在带有最少冗余的同时保证最大的连零系数的出现概率,在 DCT 图像编码中,可以对变换系数采用 Z 字形扫描。

4) 混合编码

混合编码是指综合了熵编码、变换编码或预测编码的编码方法,如 JPEG 标准和 MPEG 标准等。通过混合编码,可以综合不同编码方法的优势。

3. 第二代编码技术

1) 分形编码

分形编码是在数学家 Mandelbort 建立的分型几何理论的基础上发展起来的一种编码方法。分型编码最大限度地利用了图像在空间域上的自相似性,通过消除图像的几何冗余来压缩数据。M. Barnsley 将迭代函数系统(IFS)用于描述图像的自相似性,并将其用于图像编码。

对分形定义的一般描述如下:

① 分形应有精细的结构,有任意小比例的细节。

② 非常不规则,以至于其局部和整体都不能用传统的几何语言来描述。

③ 分形通常有某种自相似的形式,可能是近似的或是统计的。

④ 其"分形维数"一般大于其拓扑维数,并且通常能以非常简单的方法定义,由迭代方法产生。

分形编码的方法是利用图形处理技术,如颜色分割、边缘检测、频谱分析等将原始图像分割成若干子图像,然后为每个子图像寻找迭代函数,子图像以迭代函数的形式存储。

由于这样的迭代函数一般只需要几个数据表示即可，所以分形压缩可以达到较高的压缩比。

分形编码是一种新颖、独特的压缩方法。它充分考虑了自然景物的特点。其优点是：压缩比取决于图像分割后所产生的子块的大小，子块取得越大，压缩比越高；由于分形变换可把图像划分成大得多、形状复杂得多的分区，故压缩比不受分辨率的影响。其缺点有：分形编码是非对称的，压缩时计算量较大，所需时间较长，但解压缩速度很快；随着被压缩图像增大，运算量增长过快。

2）模型编码

基于模型的图像编码技术是近几年发展起来的一种很有前途的低比特率编码方法。它利用了计算机视觉和计算机图形学中的方法和理论，其基本出发点是在编、解码两端分别建立起相同的模型，针对输入的图像提取模型参数，然后根据模型参数重建图像。模型编码方法的核心是建模和提取模型参数，其中模型的选取、描述和建立是决定模型编码质量的关键因素。为了对图像数据建模，一般要求对输入图像要有某些先验的知识。根据使用的模型的不同，模型编码可以分为语义基编码和物体基编码。

基于模型的图像编码方法是利用先验模型来抽取图像中的主要信息，并以模型参数的形式表示它们，因此可以获得很高的压缩比。然而在模型编码方法的研究中还存在很多问题，例如：

① 模型法需要先验知识，不适合一般的应用。

② 对不同的应用所建模型是不一样的。

③ 在线框模型中控制点的个数不易确定，还未找到有效的方法能根据图像内容来选取。

④ 由于利用模型法压缩后复原图像的大部分是用图形学的方法产生的，因此看起来不够自然。

⑤ 传统的误差评估准则不适合用于对模型编码的评价。

3）小波变换编码

小波变换编码是随着小波变换理论的研究而提出的一种编码方式。小波变换的本质是多分辨率或多尺度地分析信号，非常适合视觉系统对频率感知的对数特性，因此，它很适合于图像信号的处理。

小波变换编码一方面具有传统编码方法的一些优点，能够很好地消除统计冗余，另一方面它的多分辨率特性充分利用了人眼的视觉特性，而且变换后的图像数据能够保持原图像在各种分辨率下的精细结构，为进一步去除其他形式的冗余提供了便利。

小波变换编码的核心问题是要对子带图像进行小波系数的量化和编码。低频子带图像包含原图像的大部分能量，即包含图像的基本特性。它在图像重构算法中起主导作用，对重建图像的质量有很大影响，因此这部分信号应精确保留。

高频子图像的系数分布符合广义高斯分布，对其系数进行粗量化编码较为有效。这也完全符合人的视觉特性，根据对人眼视觉系统的特性可知，人眼视觉灵敏度具有明显的低通特性，而且对不同方向上的敏感度也不一样，尤其是对倾斜方向的刺激不太敏感，如人眼对对角线方向子图像系数误差敏感度较低，因此可对对角线方向子图像进行粗量

化高压缩。

小波变换后的能量主要集中在低频系数分量，而其他高频系数分量大多为零值，这为高倍率压缩提供了可能。通过选择合适的具有平滑特性小波基，就可消除重建图像中出现的方块效应，减小量化噪声，获得较好的重建图像质量。用小波分析方法对图像进行编码时，主要涉及三个方面的问题：图像边界的扩展、小波基的选取和小波系数的组织。波变换编码压缩方法可分为如下两大类：基于传统的图像编码方法和基于分形理论的小波变换图像编码方法；基于传统的图像编码方法，包括零树小波编码、基于塔式网络矢量量化的小波变换编码、基于 LBG 算法的小波变换编码、基于标量量化的小波变换编码等。

针对分形图像编码尚存在的缺点，如编码算法耗时、自然图像不一定具有严格的分形结构而无法达到预期的高压缩比、高压缩倍率时的方块效应等，有人提出了基于小波变换的分形编码。它具有以下特点：

① 采用平滑小波可去除传统分形变换中存在的方块效应。

② 小波表示使图像的四叉树分割十分自然。

③ 可将零树算法看成是该算法的一个特例。图像经过金字塔形离散小波变换后的系数在小波域内可组成分层树状数据结构小波树。这些跨越不同分辨率的小波树之间存在一定的相似性，可通过分形变换来描述。基于小波变换的分形压缩过程就是一个由分层树状结构的顶部开始一层层地向下预测其余系统的过程，而这个由上至下、由粗至细的预测过程是通过分形编码来实现的。

目前，小波变换编码已获得了较好的编码效果，是现代图像压缩技术研究的热点之一，也是一种十分有前途的方法。

7.1.3 数字图像处理技术

1. 数字图像处理的基本概况

20 世纪 20 年代，图像处理首次应用于改善伦敦和纽约之间海底电缆发送的图片质量。到 20 世纪 50 年代，数字计算机发展到一定的水平后，数字图像处理才真正引起人们的兴趣。1964 年美国喷气推进实验室用计算机对“徘徊者七号”太空船发回的大批月球照片进行处理，收到了明显的效果。20 世纪 60 年代末，数字图像处理具备了比较完整的体系，形成了一门新兴的学科。20 世纪 70 年代，数字图像处理技术得到迅猛的发展，理论和方法进一步完善，应用范围更加广泛。在这一时期，图像处理主要和模式识别及图像理解系统的研究相联系，如文字识别、医学图像处理、遥感图像的处理等。从 20 世纪 70 年代后期到现在，各个应用领域对数字图像处理提出越来越高的要求，促进了这门学科向更高级的方向发展。特别是在景物理解和计算机视觉(即机器视觉)方面，图像处理已由二维处理发展到三维理解或解释。近年来，随着计算机和其他各相关领域的迅速发展，例如在图像表现、科学计算可视化、多媒体计算技术等方面的发展，数字图像处理已从一个专门的研究领域变成了科学研究和人机界面中的一种普遍应用的工具。

图像处理技术基本可以分成两大类：模拟图像处理和数字图像处理。数字图像处理

一般都用计算机处理或实时的硬件处理，因此也称之为计算机图像处理，由于它处理精度高，包含信息量大，广泛应用于空间探测、遥感、生物医学、人工智能以及工业检测等许多领域。数字图像处理技术主要包括如下内容：几何处理、算术处理、图像增强、图像复原、图像重建、图像编码、图像识别、图像理解。数字图像处理技术的发展涉及信息科学、计算机科学、数学、物理学以及生物学等学科，因此数理及相关的边缘学科对图像处理科学的发展有越来越大的影响。近年来，数字图像处理技术日趋成熟并促使这些学科也产生了新的发展。下面将简单介绍数字图像处理学的发展和现状。

2. 数字图像处理的常用方法

1）图像变换

由于图像阵列很大，直接在空间域中进行处理，涉及计算量很大。因此，往往采用各种图像变换的方法，如傅里叶变换、沃尔什变换、离散余弦变换等间接处理技术，将空间域的处理转换为变换域处理，不仅可减少计算量，而且可获得更有效的处理（如傅里叶变换可在频域中进行数字滤波处理）。目前新兴研究的小波变换在时域和频域中都具有良好的局部化特性，它在图像处理中也有着广泛而有效的应用。

2）图像编码压缩

图像处理中另一至关重要的问题是图像数据的压缩。特别是在获取了大量的静态和动态图像后，要将其传输到用户终端或存储图像以备今后使用时，遇到的最大困难就是图像巨大的数据量。因此，图像信息的压缩在图像的存储、传递，乃至后面谈到的多媒体技术中都是至关重要的问题。

图像编码压缩技术可减少描述图像的数据量（即比特数），以便节省图像传输、处理时间和减少所占用的存储器容量。压缩可以在不失真的前提下获得，也可以在允许的失真条件下进行。编码是压缩技术中最重要的方法，它在图像处理技术中是发展最早且比较成熟的技术。图像压缩编码的研究有比较久远的历史，直至目前，仍在不断探索新的技术和方法。图像压缩编码的方法，主要是消除图像存储过程中产生的大量数据冗余。为了得到较好的结果，可用预测编码、变换编码、熵编码等高清晰度图像压缩编码方法。

3）图像增强和复原

图像增强和复原的目的是为了提高图像的质量，如去除噪声，提高图像的清晰度等。图像增强不考虑图像降质的原因，突出图像中所感兴趣的部分。如强化图像高频分量，可使图像中物体轮廓清晰，细节明显；如强化低频分量可减少图像中噪声影响。图像复原要求对图像降质的原因有一定的了解，一般来说应根据降质过程建立“降质模型”，再采用某种滤波方法，恢复或重建原来的图像。

获取到的图像，通常带有各种畸变和干扰。例如有成像器件的缺陷。如带宽限制造成图像模糊、成像过程中不可避免的热噪声和其他干扰源带来的各种干扰噪声等，为了获取为人们观测处理所需要的良好质量的图像，需要引入图像处理。这包括图像的增强和图像的复原。图像的增强是采用增强轮廓边缘，进行灰度和颜色等变换，使图像更适合于人们观测和处理的需要。而图像的复原则是为消除或减小图像获取和传输过程中

造成的图像的损伤和退化，这包括图像的模糊、图像的干扰和噪声等，尽可能获得原来真实的图像。图像复原往往是比较困难和复杂的逆滤波过程。尤其在造成图像退化的过程比较复杂和难以预测时，图像复原就更难了。不论图像的增强或复原，都必须对整幅图像的所有像素进行运算，出于图像像素的数量巨大，其运算量也是很大的。

4）图像分割

图像分割是数字图像处理中的关键技术之一。图像分割是将图像中有意义的特征部分提取出来，其有意义的特征包括图像的边缘和区域等，这是进一步进行图像识别、分析和理解的基础。虽然目前已提出不少边缘提取、区域分割的方法，但还没有一种能普遍适用于各种图像。因此，对图像分割的研究还在不断深入之中，是目前图像处理中研究的热点之一。下面介绍几种常见的图像分割方法。

基于阈值的分割方法是一种直接对图像灰度信息阈值化处理的分割算法，就是简单地用一个或几个阈值将图像灰度直方图进行分类，将灰度值在同一个灰度类内的像素归为同一个物体，直接利用图像的灰度特性进行分割。因此有实现简单、成本低廉、实用性强等优点；但是当图像中灰度差异不明显、或者各物体的灰度范围值有大部分重叠现象时，往往难以得到准确的分割结果，从而产生很多过分割错误。

基于边缘的分割方法利用了边缘总是以强度突变的形式出现的特性，或者说不同区域之间像素灰度值变化比较剧烈的特点；根据相关的数学知识，这类方法一般采用图像一阶导数极值和二阶导数过零点信息作为边缘点的判断依据，边缘定位准确，运算速度快。但边缘的连续性和封闭性难以保证，对于复杂图像的分割效果较差，如可能出现边缘模糊、边缘丢失等现象。边缘检测方法常常依赖于边缘检测算子，从而找到图像边缘；常用的检测算子有：Roberts 算子、Sobel 算子、Prewitt 算子、Canny 算子、Laplacian 算子和 Marr 算子（即 LOG 算子）。

基于区域的图像分割考虑了图像的空间信息，如图像灰度、纹理、颜色和像素统计特性等，进而将目标对象划分为同一区域的分割方法。常见的区域分割方法有：区域生长法、分裂合并法和分水岭分割方法。

5）图像描述

图像描述是图像识别和理解的必要前提。作为最简单的二值图像可采用其几何特性描述物体的特性，一般图像的描述方法采用二维形状描述，它有边界描述和区域描述两类方法。对于特殊的纹理图像可采用二维纹理特征描述。随着图像处理研究的深入发展，已经开始进行三维物体描述的研究，提出了体积描述、表面描述、广义圆柱体描述等方法。

6）图像分类（识别）

图像分类（识别）属于模式识别的范畴，其主要内容是图像经过某些预处理（增强、复原、压缩）后，进行图像分割和特征提取，从而进行判决分类。图像分类常采用经典的模式识别方法，有统计模式分类和句法（结构）模式分类，近年来新发展起来的模糊模式识别和人工神经网络模式分类在图像识别中也越来越受到重视。

7.1.4 数字图像内容安全的技术分类

数字图像信息安全，是伴随着计算机网络和多媒体技术的迅速发展而产生的新问题，如何保证数字图像信息的安全已成为国际上的热门研究课题。常见的数字图像内容安全的技术有以下几种。

1. 图像加密技术

通过图像加密操作后，原来的数字图像变为类似于信道随机噪声的信息，这些信息对不知道密钥的网络窃听者是不可识别的(除非进行了有效破译)，进而可以有效地保护传输中的图像数据。随着人们对知识产权的重视及娱乐工业的发展，可以预见，图像加密技术会有广阔的应用前景。

2. 数字图像水印技术

数字图像水印技术是利用数字图像中普遍存在的冗余数据和随机性把版权信息嵌入在数字图像中从而起到保护数字图像版权或完整性的一种技术。作为版权信息嵌入到数字图像中的秘密信息即称为数字水印(digital watermark)，它可以是无意义的随机序列，也可以是文字、图像、声音等有意义的信息。

由于数字图像水印技术的目的在于保证水印数据不被侵犯和发现，同时还必须考虑水印数据在经历各种环境、正常和非正常数据操作之后是否仍具有免遭破坏的能力，因此，为使数字图像水印技术得以实施，它必须具备下面的特性：

1）透明性

对于以模拟方式存储和分发的信息(如电视节目)，或是以物理形式存储的信息(如报刊、杂志)，用可见的标志就足以表明其所有权。但在数字方式下，标志信息极易被修改或擦除。因此应根据多媒体信息的类型和几何特性，利用用户提供的密钥将水印隐藏到一系列随机产生的位置中，使人无法察觉。

2）鲁棒性

鲁棒性是指数字图像经过一些处理、数字图像数据发生一定程度的变化后，版权所有者仍然可以证明水印的存在。可能的处理包括：

① 几何变形：对图像进行尺寸缩放、剪裁、扭转等。

② 有损压缩：常用的图形文件格式 JPEG 就属于有损压缩。它先将图像用 DCT 函数转换到频率域，然后对其量化，在量化过程中忽略掉一些感知上不重要的成分，以达到压缩文件尺寸的目的。虽然肉眼看不出来，但压缩后图像的精度肯定有所降低。

③ 信号处理：如调整图像和视频的对比度、亮度、色度，以及模/数、数/模转换等。

3）不可检测性

水印作品和普通作品在统计噪声分布上不存在区别，攻击者无法用统计学方法确定水印的位置。

4）安全性

数字图像水印技术应具有较强的抗攻击能力，能够承受一定程度的人为攻击，而暗

藏的水印不被破坏。

5）自恢复性

由于经过一些操作或变换后，可能会使原图产生较大的破坏，如果只根据留下的片段数据，仍能恢复隐藏信号，而且恢复过程不需要宿主信号，这就是所谓的自恢复性。

3. 数字图像隐写分析技术

数字图像隐密分析技术主要有三个阶段：首先，需要进行判断图像中是否存在秘密信息的一般性隐密分析；其次，在用一般性隐密分析判断出图像中是否含有秘密信息基础上，需要使用针对性隐密分析方法判断可能使用的隐密方法，并确定出被嵌秘密信息长度和嵌入位置；最后，确定隐藏方法、隐密软件和嵌入密钥来提取秘密信息以作为确凿证据。

7.2 数字图像内容加密技术

在网络上传输图像数据，很多情况下要求发送方和接收方在保密的情况下进行，如军用卫星所拍摄的图片、军用设施图纸、新型武器图、金融机构的建筑图纸等；还有些图像信息，如在远程医疗系统中，医院中患者的病历(其中包括患者的图像)，根据法律必须要在网络上加密后方可传输。不言而喻，由 Internet 传输图像数据不但方便快捷，不受地域限制，而且省时省力，节约开支，提高效率。但由于某些图像数据的特殊性，即发送和接收的双方都不希望网络上所传输的图像数据被未授权者所浏览或处理，因为这些图像信息不但涉及个人隐私，而且有的涉及国家安全，因而图像数据的保护越来越受到社会的普遍重视。

7.2.1 数字图像加密技术分类

数字图像加密就是在发送端采用一定的算法作用于一幅图像明文，使其合成不可识别的密文，达到图像保密的目的。在接收端采用相应的算法解密，恢复出原文。其通用算法模型如图 7-2 所示。



图 7-2 数字图像加密通用模型

数字图像加密有多种分类方法，如果按照加密手段的不同，可分为：基于现代密码体制的加密方法、基于混沌理论的加密方法、基于矩阵变换或像素变换的加密方法等。按照加密对象的不同，可分为：对空间域像素值的加密方法、对变换域系数的加密方法等。按照加密时结合的技术的不同可分为：结合图像编码技术的加密方法、结合图像压缩技术的加密方法、结合神经网络的加密方法等。

这些方法既相互独立又相互关联，甚至一些方法的结合使用能达到意想不到的加密效果。在不同的应用场合、不同的加密要求下，可以选择适当的加密方法。下面着重介绍其中几种典型加密方法的原理和优缺点。

7.2.2 典型的数字图像加密算法

1. 基于矩阵变换及像素置换的图像置乱加密技术

图像置乱加密技术的基本方法是把一幅图像经过变换或利用数学上的知识，搅乱像素位置或颜色，将原来有意义的图像信息变换成一幅“杂乱无章”的图像，无法辨认出原始图像信息，从而达到在一定程度上迷惑第三方的目的。为了确保其机密性，算法中一般引入密钥。图像合法接受方借助密钥，通过相应算法的逆变换可解密出原始图像，这一过程又称去乱。

目前，数字图像置乱加密的方法已有许多种，这些方法在一定的应用范围中各自起到了积极的作用。由于置乱加密不仅用于图像信息的保密，同时也是图像信息隐藏、图像信息分存、数字水印技术等工作的基础，因此置乱加密算法的优劣也直接影响到其他处理的效果。

1）Arnold 变换

设像素的坐标 $x,y\in S=\{0,1,2,\cdots,N-1\}$，Arnold 变换为：

$$\begin{bmatrix} x' \\ y' \end{bmatrix}=\begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix}(\bmod N),\quad x,y\in S \tag{7-4}$$

记变换中的矩阵为 $\boldsymbol{A}$，反复进行这一变换，则有迭代公式：

$$\boldsymbol{Q}_{ij}^{n+1}=\boldsymbol{A}\boldsymbol{Q}_{ij}^{n}(\bmod N),\quad n=0,1,2,\cdots \tag{7-5}$$

其中，$\boldsymbol{Q}_{ij}^{0}\in S$，$\boldsymbol{Q}_{ij}^{n}=(i,j)^{\mathrm{T}}$ 为迭代第 n 步时点的位置。

Arnold 变换可以看作是裁剪和拼接的过程。通过这一过程将离散化的数字图像矩阵中的点重新排列。由于离散数字图像是有限点集，这种反复变换的结果，在开始阶段 s 中像素点的位置变化会出现相当程度的混乱，但由于动力系统固有的特性，在迭代进行到一定步数时会恢复到原来的位置，即变换具有庞加莱回复性。这样，只要知道加密算法，按照密文空间的任意一个状态来进行迭代，都会在有限步内恢复出明文（即要传输的原图像）。这种攻击对于现代的计算机来说其计算时间是很短的，因而其保密性不高。

2）其他置乱加密技术

相对位置空间而言，基于色彩空间的置乱加密技术是指通过数学知识或其他性质，置乱原始图像像素的灰度值或色彩值，同样可起到扰乱原图信息的目的。例如基于灰度变换的置乱加密方法，其思想来源于数字图像处理中的灰度直方图变换，置乱加密算法中的密钥增加了破解的难度；可采用密码学加密算法对图像灰度进行变换，研究空间更广泛，算法运行时间较短。人们意识到置乱加密技术不仅可以考虑将图像的像素位置置乱，像素灰度值也可以进行置乱处理。后来，有两种新的置乱变换被提出：准逆序置乱和准抖动置乱，这是针对数字图像灰度空间中两种变换的置乱加密。在图像信息隐蔽存储与传输中，这类图像变换具有重大的应用价值。

混沌系统在一定的控制参数范围内会出现混沌现象,产生的混沌序列具有确定性、伪随机性、非周期性和不收敛等性质,并且对初始值有极其敏感的依赖性。由于混沌天然的优势,人们多引用 Logistic 映射产生实数值混沌,采用不同的量化方法对其量化为混沌序列,然后运用到图像置乱加密中来,加密效果非常好,再结合一定的其他算法,可以达到快速、安全性高的效果。不可否认,混沌的引入为图像置乱加密带来了又一新的发展方向。

基于变换空间的置乱加密也是图像置乱加密中的又一新领域。它主要是指对数字图像的变换域(如离散余弦变换 DCT、离散傅里叶变换 DFT、小波变换等)的系数进行置乱,扰乱图像信息。不过较成熟的变换域置乱加密算法还有待进一步研究和开发。

2. 基于现代密码体制的图像加密技术

Claude Shannon 于 1949 发表了一篇题为"保密系统的信息理论"的文章,用信息论的观点对信息保密问题做了全面的阐述,建立了现代密码学理论。对于图像数据来说,这种加密技术就是把待传输的图像看做明文,通过各种加密算法,如 DES、AES 等,在密钥的控制下,达到图像数据的保密通信。这种加密机制的设计思想是加密算法可以公开,通信的保密性完全依赖于密钥的保密性(即满足 Kerckhoffs 假设)。其原理框图如图 7-3 所示。

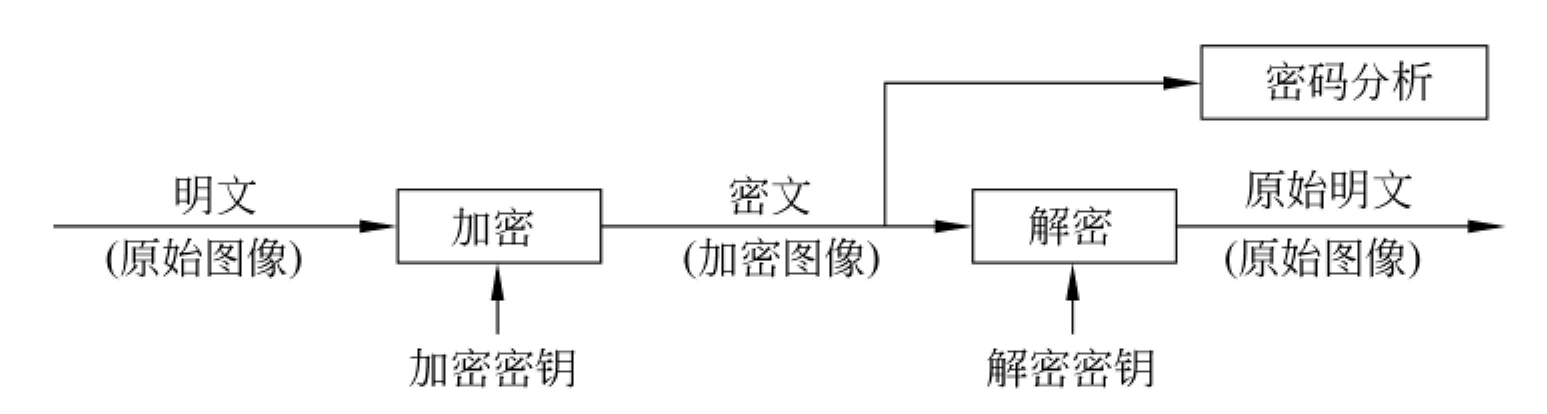


图 7-3 密钥控制下的保密通信框图

图 7-3 中,加密密钥和解密密钥可以相同也可以不同,并依此来划分出两种基本的密码算法,即对称算法和非对称算法(也叫公开密钥算法)。

由于数字图像的数据量通常较大,若直接采用现代密码体制中的标准算法进行加密,其处理效率通常较低。

3. 基于混沌的图像加密技术

基于混沌的图像加密技术是近年才发展起来的一种新型密码技术。它是把待加密的图像信息看作是按照某种编码方式的二进制的数据流,利用混沌信号来对图像数据流进行加密。混沌之所以适合于图像加密,这是与它自身的动力学特点密切相关的。

混沌加密的原理就是在发送端把待传输的有用信号叠加(或某种调制机制)上一个(或多个)混沌信号,使得在传输信道上的信号具有类似随机噪声的性态,进而达到保密通信的目的。在接收端通过对叠加的混沌信号去掩盖(或相应的解调机制),去除混沌信号,恢复真正传输的信号。

混沌加密方法属于对称加密体制的范畴,这种加密体制的安全性取决于密钥流发生

器(即混沌)所产生的信号与随机数的近似程度,密钥流越接近随机数,其安全性越高,反之则容易被攻破。混沌加密方法是符合现代密码学要求的,其近阶段的主要研究方向是寻找更加随机的混沌流,并解决混沌流的同步问题。

4. 基于秘密分割与秘密共享的图像加密技术

秘密分割就是把消息分割成许多碎片,传一个碎片本身并不代表什么,但把这些碎片放到一起消息就会重现。这种思想用于图像数据的加密就是在发送端先要把图像数据按某种算法进行分割,并把分割后的图像数据交给不同的人来保存;而在接收端需要保存秘密的人的共同参与才能恢复出原始待传输的图像数据。为了实现在多个人中分割一幅秘密图像信息,可以将此图像信息与多个随机位异或成“混合物”。例如 Trent 可将一幅图像信息划分为 4 部分并按如下协议实现:

- Trent 产生 3 个随机位串 R,S,T,每个随机位串和图像信息 M 一样长。
- Trent 用这 3 个随机位串和 M 异或得到 U:$M \oplus R \oplus S \oplus T = U$。
- Trent 将 R 给 Alice,S 给 Bob,T 给 Carol,U 给 Dave。

Alice、Bob、Carol、Dave 在一起可以重构待传输的秘密图像信息,$M \oplus R \oplus S \oplus T = M$。

在这个协议中,Trent 作为仲裁人具有绝对的权利,他知道秘密的全部,他可以把毫无意义的东西分发给某个人,并宣布是秘密的有效部分,并在秘密恢复之前没有人知道这是不是一句谎话(他可以把“秘密”分发给 Alice、Bob、Carol、Dave 四个人,并宣布秘密都是有效的,但实际上只需要 Alice、Bob、Carol 三人就可恢复秘密)。

这个协议存在这样一个问题:如果秘密的一部分丢失了而 Trent 又不在,就等于把秘密丢失了,而且这种一次一密的加密体制是有任何计算能力和资源的个人和部门都无法恢复秘密的。

5. 基于压缩编码技术的加密方法

数字图像的大数据量是图像的一个显著特点,在数字图像处理研究中,图像的压缩编码技术格外引人注目。许多学者将二者有机地结合在一起,取得了令人瞩目的成绩,丰富了图像加密技术。

6. 基于变换域的加密方法

针对数字图像数据的特点,人们对其加密方法的思想不单单限制在图像的像素空间域上,而将更多的目光投向了图像的变换域。利用传统的加密方法对图像文件加密时,只是对图像全部数据进行加密,但对大数据量的图像数据进行加密,显然不太现实。

图像的变换域是相对于图像的像素空间域而言的,一般地可以利用 DCT、快速傅里叶变换(Fast Fourier Transform,FFT)以及小波变换等方法来实现图像空间域和变换域之间的转换。基于变换域的加密方法主要是将图像作变换后,对变换系数进行保密处理,这样大大减少了保密数据,提高了加密效率,同时也增加了转化时间。

7.3 数字图像内容隐写与水印技术

数字水印技术是将一些标识信息直接嵌入数字载体当中(包括多媒体、文档等)或是间接表示(修改特定区域的结构),它不影响原载体的使用价值,也不容易被探知和再次修改,但可以被生产方识别和辨认。通过这些隐藏在载体中的信息,可以达到确认内容创建者、购买者、传送隐秘信息或者判断载体是否被篡改等目的。它是实现版权保护的有效办法,是信息隐藏技术的一个重要研究方向。

数字图像水印技术是指用信号处理的方法在图像中嵌入隐含标记。一个数字图像水印系统主要包括水印的生成、嵌入和检测三个部分。图 7-4 展示了整个模型的框图。

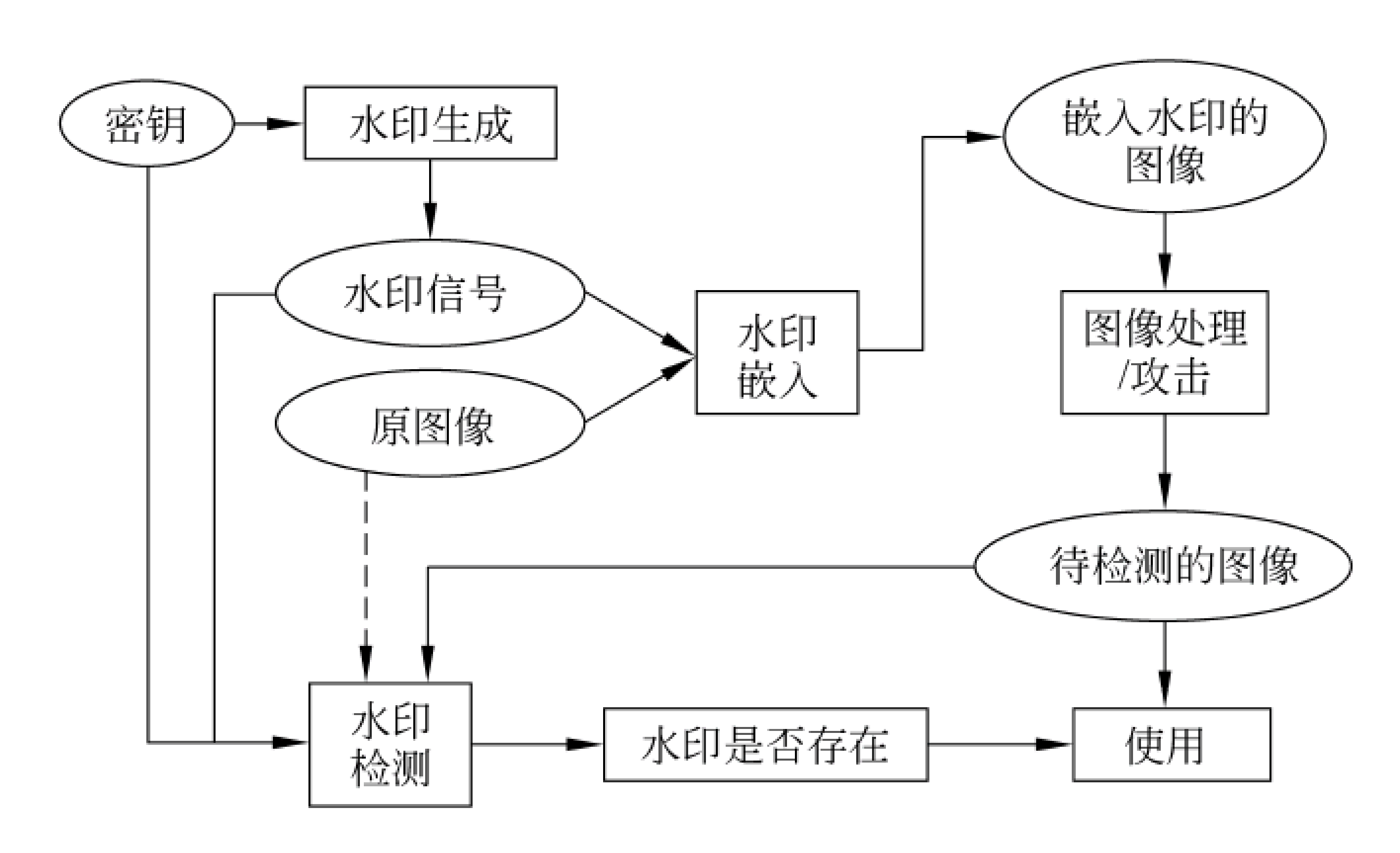


图 7-4 一个数字图像水印系统的模型

图中虚线表示在某些情形(如盲水印系统)中,原图像是不需要的。

生成数字水印是完成向数字图像中嵌入水印的最关键的一步。若 m 表示原始信息,F 表示水印生成函数,X 表示要嵌入水印的原始图像,K 表示密钥,W 表示要加入的水印,则有:

$$W = F(m, X, K)$$

一般情况下,数字水印的生成过程如图 7-5 所示。



图 7-5 水印生成过程

水印的嵌入就是把生成的水印信息进行适当变换嵌入到数字图像中的过程。水印嵌入模型如图 7-6 所示,其输入信号是水印信息、载体数据和一个可选的密钥,水印系统通常使用密钥来确保安全。水印可以是随机数字序列、图像或文本的任意形式信息。常用的嵌入公式有:

$$v_i^w = v_i + \alpha w_i \tag{7-6}$$

$$v_i^w = v_i(1 + \alpha w_i) \tag{7-7}$$

其中，v_i、v_i^w 分别表示图像像素和嵌入水印后的图像像素；w_i 为水印信号分量；参数 α 为小于 1 的水印嵌入强度因子，α 的选择必须考虑图像的性质和视觉系统的特性，在保证水印不可见的前提下，尽可能提高嵌入水印的强度。

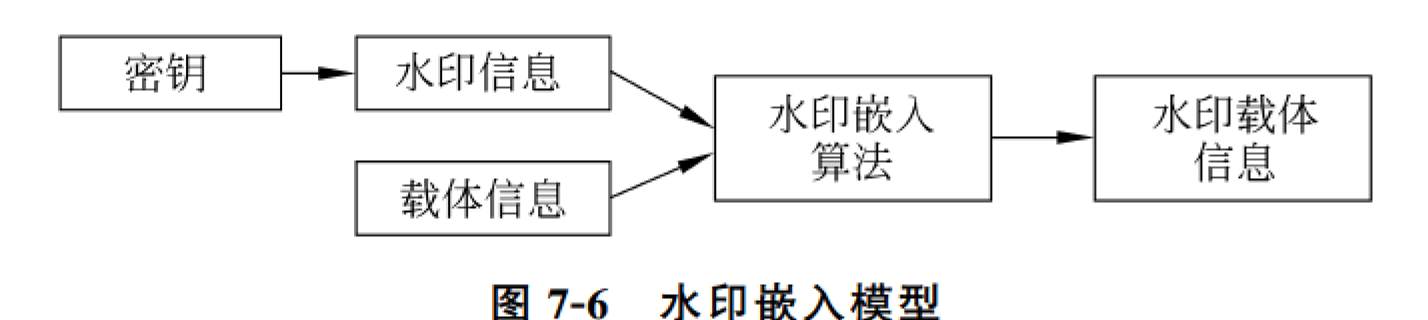


图 7-6 水印嵌入模型

数字水印的检测(提取)是数字水印的关键技术之一。一般情况下，在提取水印之前应先检测水印是否存在，然后根据检测(提取)密钥，采用嵌入算法的逆算法，检测(提取)待证实的每个水印。若用 G 表示水印检测函数，X^* 为待检测图像，W^* 为待证实的水印，X 为原始载体图像，K 为密钥，则有：

$$W^* = G(X^*, X, K) \tag{7-8}$$

水印检测模型如图 7-7 所示。

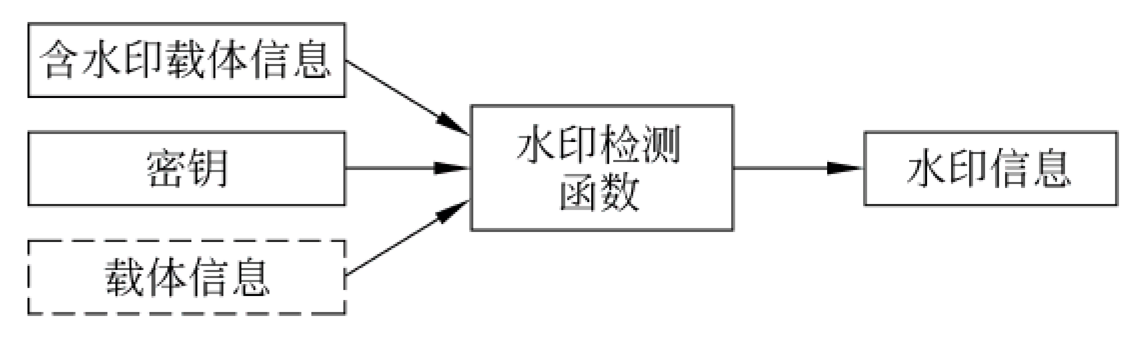


图 7-7 水印检测模型

7.3.1 数字图像水印的分类

数字水印的分类方法有多种。从加水印后图像中的水印是否可见可分为可见水印和不可见水印两大类(大部分情况下，水印是不可见的)。

从水印的来源可分为独立于图像的水印和图像自适应的水印。独立于图像的水印可以是随机产生的也可以是事先给定的，而图像自适应的水印是利用原始图像的特性生成的水印。

从加水印图像的抗过滤或压缩等能力来分，可以分为脆弱水印、半脆弱水印和鲁棒水印。脆弱水印对任何图像变换或处理都非常敏感，半脆弱水印对某些特定的图像处理方法有鲁棒性而对其他的处理不具备鲁棒性，鲁棒水印对常见的各种图像处理方法都具有鲁棒性。

从水印检测是否需要原始图像参与来分，可以分为私有水印和公有水印。私有水印的检测需要原始图像参与，公有水印的检测不需要原始图像的参与。因此私有水印对原始图像的依赖性比较强，这在网络上是很不利的，而公有水印则只依赖于图像本身，不需要原始图像。

对于数字图像来说，水印技术就是通过改变图像数据的值来加入水印，根据加入位

置的不同，水印技术又可分成时域法和频域法。时域法直接改变图像元素的值，一般是在图像元素的亮度或色度中加入调制的水印信号。频域法利用某种数学变换，将图像用频域表示，通过改变图像的某些频域系数来加入水印，然后利用反变换生成含水印的图像，常用的数学变换有离散傅里叶和离散余弦变换等。

7.3.2 典型的数字图像水印算法

近年来，在图像水印方面已提出了很多算法，比较典型的有空域数字水印、变换域数字水印、压缩域数字水印、NEC 数字水印、生理模型数字水印等。下面就分别对上述方法进行介绍。

1. 空间域数字水印算法

首先生成一个 M 序列形式的水印，然后将图像的像素比特位压缩，把水印直接替换图像像素的最低位(Least Significant Bit，LSB)或者采用线性叠加的方法将水印嵌入到 LSB 上。这种水印具有一定的鲁棒性，但由于它位于图像的 LSB 上，所以很容易被去除。

G. Voyatzis 和 I. Pitas 提出了一种基于混沌系统的算法，即先设计一个基于 Torus Automorphism 的混沌系统，将一幅用作水印的 $N\times N$ 图像 S 和密钥 K 输入到这个混沌系统中，得到一个混沌的水印图像 S'。然后，在待嵌水印的 $M_1\times M_2$ 图像 I_O 中，选择一块 $N\times N$ 区域，以一定的方法叠加嵌入水印 S'。这种方法具有很强的鲁棒性，能较好地抵御几何攻击。

2. 变换域算法

I. J. Cox 等提出了基于 DCT 域的扩频水印算法。先用密钥生成一个长度为 1000 的服从正态分布 $N(0,1)$的伪随机序列，然后对图像进行全局二维 DCT 变换，选取最大的 1000 个交流(AC)系数，采用一个合适的嵌入公式将水印嵌入。例如：

$$x_i' = x_i(1+\alpha w_i) \tag{7-9}$$

式中：x_i 为第 i 个最大的 DCT 交流系数，x_i'为嵌入水印后的系数，w_i 为第 i 个水印分量，α 是个常量，表示水印的嵌入强度。最后做一个反 DCT 变换，得到嵌有水印的图像。

这种水印具有很强的鲁棒性，对串谋攻击、JPEG 压缩、缩放、剪切、重复加水印等多种攻击方式均能较好地抵挡。Cox 等此后又做了大量工作，对这一算法进行了改进。

基于其他变换域，如 DFT 等算法也有很多，但主要原理是一致的，即将图像从空域变换到某个适合处理的变换域，再修改该域中最重要的若干个系数以嵌入水印。如有人提出了一种在 DFT 域中嵌入一个对称环形水印的算法。另外有方法则先生成一个具有空间自相似特性的水印，再将图像以 Harr 小波基函数分解为 4 级，将水印嵌入到最高级和次高级的细节区中。这两种算法不仅对滤波和 JPEG 压缩具有很强的鲁棒性，而且能有效抵抗几何攻击。

3. 生理模型算法

人的生理模型包括人类听觉系统(Human Auditory System,HAS)和人类视觉系统HVS(Human Visual System,HVS)。该模型不仅被多媒体数据压缩系统利用,同样也可以供数字图像水印系统利用。利用视觉模型的基本思想均是利用从视觉模型导出的JND(Just Noticeable Difference)描述来确定在图像或声音的各个部分所能容忍的数字水印信号的最大强度,从而能避免破坏视觉或者听觉的质量。也就是说,利用生理模型来确定与数据相关的调制掩模,然后再利用其来嵌入水印。这一方法具有较好的透明性和强健性。

4. 其他算法

1) 分形水印

基于图像分形压缩的分形水印是由Puate和Jordan提出的。令嵌入的信息为b,$b\in\{0,1\}$,在图像中随机选取一区域块,将它分成两个相等的子区域块,给每一个子块分配1比特信息,然后进行搜索,只有当子块中含有相对应比特值时,该区域块才会被编码。

在恢复过程中,先对含水印图像作分形压缩,然后进行全局搜索,被标记块的位置即包含了嵌入信息。实验表明这种水印可以有效抵抗JPEG压缩,缺点是计算量大、速度慢,这主要是由分形压缩造成的。

2) 基于特征的水印算法

1999年Kutter等最先提出第二代水印的概念,建议水印嵌入在感知有意义的特征区域中。对于图像来说,可能是边缘、拐角和纹理区域,或者是突出点所在的区域。Bas等提出了基于图像特征点的水印方案,首先提取特征点,然后将水印嵌入在图像特征点组成的三角形网格中。此外还有局部化数字水印算法,该算法利用图像中相对稳定的特征点来标示水印的嵌入位置,并在每个特征点的局部区域内独立地嵌入水印。这样,当只剩部分图像时,仍能够通过这些特征点来定位并提取水印。

7.4 数字图像内容隐写分析技术

7.4.1 数字图像隐写分析技术分类

隐写术是利用人的感觉器官对数字信号的感觉冗余,通过一定的算法将隐密信息嵌入到数字载体(包括声音、图像、视频等)中,以不被人的知觉系统所觉察,从而实现隐蔽通信。隐写分析作为隐写术的对立面,是指对隐写术的检测和攻击,对可疑的载体信息进行攻击,以实现隐密信息的检测、破坏,甚至提取隐密信息。隐写术通用的隐写过程可表示如下:

$$S' = S + f(S,M)$$

式中,S和S'分别代表载体消息和嵌入秘密消息后的隐藏消息,M为待嵌入的秘密消息。

隐写分析的过程就是从 S' 中检测出 M 甚至提取 M。

隐写术与隐写分析的一般框架可用图 7-8 的“囚犯”问题描述。秘密信息 E 的通信双方 Alice 与 Bob 应用隐写术将 E 经过密钥 K 加密后嵌入到公开的载体中，利用公开信道传输载密体，公用信道的看守者 Eve 可获得 Alice 与 Bob 之间的通信。如果只检查通信中是否含有隐蔽通信，则称 Eve 是一个被动看守者，如果主动去修改获得的通信，甚至假冒通信的一方伪造秘密信息并传给通信的另一方，则称 Eve 是一个主动看守者。根据这样的通信框架，隐写分析可分为检测、提取、混淆、还原等层次，从公开发表的文献看，目前国内外的研究重点在于检测，关于隐蔽信息的提取也开始受到人们的关注。

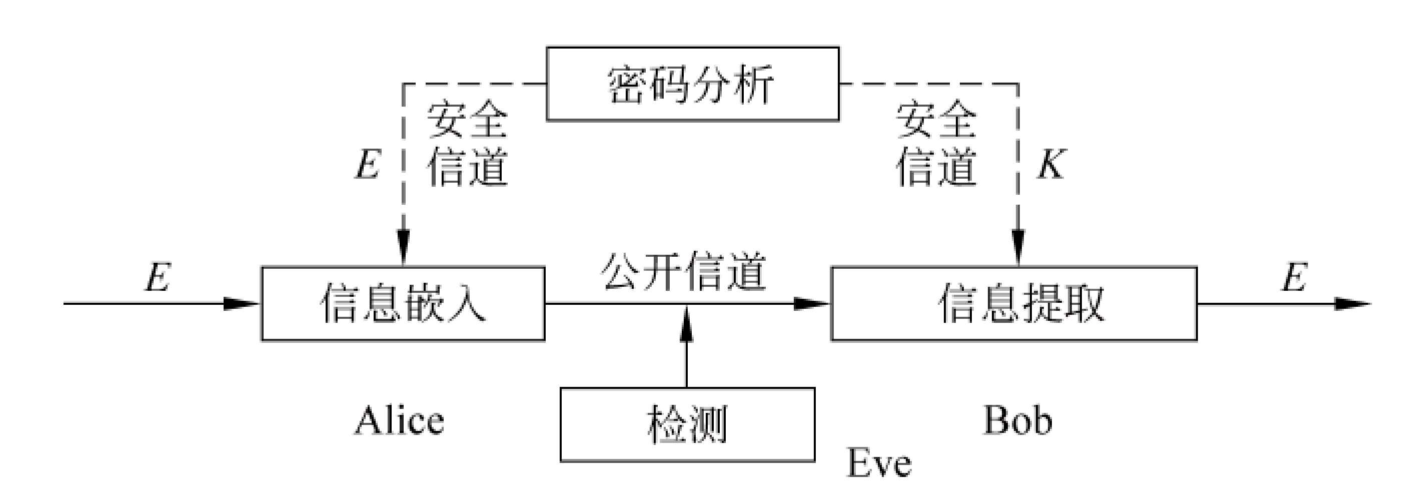


图 7-8　隐写术与隐写分析的一般框架

数字图像隐写分析方法大致可以分为专用隐写分析、通用隐写分析、针对 JPEG 图像的隐写分析方法、针对 F5 的隐写分析、基于小波特征函数统计矩的隐写分析等。其中，专用隐写分析方法根据提取特征所在的域不同可分为空域隐写分析和变换域隐写分析；通用隐写分析的攻击对象包括了空域与变换域隐写术，是隐写分析领域不可忽视的部分。下面将重点介绍各种典型方法的原理。

7.4.2　典型的数字图像隐写分析算法

1. 专用隐写分析方法

专用隐写分析方法根据提取特征所在的域不同，可分为空域隐写分析和变换域隐写分析。

1）空域隐写分析

空域法是直接改变图像元素的值，一般是在图像元素的亮度或色度中加入隐藏的内容。如 LSB 算法，它通过调整伪装载体某些像素数据的最低 1 或 2 位有效位来隐藏信息，致使所隐藏的信息靠视觉很难被发现。空域类算法的特点是只需对隐秘载体进行很小的、不易察觉的改变就能隐藏很大的信息量，计算速度较快。但从基本原理上看，该算法所隐藏的信息是极为脆弱的，若载体图像有微小的改变，隐藏信息就可能丢失。空域隐写分析的攻击对象主要是空域 LSB 隐写术，包括 EzStego、S-Tools、BPCS 等，是隐写分析研究初期非常活跃的部分。

Westfeld 等采用 Chi-square 统计量统计调色板图像嵌入秘密消息前后出现近似颜色对概率比，可以可靠地检测连续嵌入秘密消息的调色板图像，但对随机嵌入的真彩色图像检测无效。

Fridrich 等提出的 RS 检测法把图像像素分成规则类、异常类和不可使用类，根据待测图像 LSB 置换操作前后每类像素组的变化曲线可以可靠地检测灰度和真彩色图像并估计嵌入量，但算法的检测结果直接受载体图像随机性、噪声和秘密信息嵌入位置影响。

Dumitrescu 等提出的样本对分析法达到了与 RS 最优检测等效的结果。算法根据相邻像素值的奇偶性质将像素对分为 4 种基本集合，秘密消息的嵌入导致像素对从一个集合转换到另一个集合，根据集合更改的比例采用二次方程来估计嵌入量。该方法适用于对连续信号采样的检测，但检测结果直接受秘密信息嵌入位置的影响，对非随机嵌入无效。

张涛等定义了差分直方图的转移系数作为 LSB 平面与图像其余比特平面之间的弱相关性度量，并在此基础上构造了载体图像与隐藏图像的分类器。在嵌入量比较大的情况下该算法检测效果优于 RS，但检测效果受载体图分布、嵌入位置和秘密消息随机性的影响。

空域隐写分析比较多的围绕颜色对现象进行研究 ，研究的方法经历了从简单分析隐藏图像颜色对到采用比较复杂的实验手段(如再次嵌入秘密消息归类、划分集合等)来获得颜色对变化量的过程，总体来说适用性与实用性比较低。

2）变换域隐写分析

变换域隐写算法是利用某种数学变换，将图像用变换域(如频域)表示，通过更改图像的某些变换域系数加入待隐藏信息，然后再利用反变换来生成隐藏有其他信息的图像。常见的变换域算法有：基于 DCT 的变换域算法、基于 DWT 的变换域算法。变换域算法具有很好的鲁棒性(指不因图像文件的某种改动而导致隐藏信息丢失的能力)，对传输过程中的图像压缩、滤波以及噪声均有一定的抵抗力，并且很多方法还结合了当前的图像和视频压缩标准(如 JPEG、MPEG 等)，具有实际意义。变换域隐写分析的攻击对象主要是 DCT 域隐写术，包括 JSteg、F5、Outguess、MB。

Fridrich 等通过解压缩待测图像、裁剪、再压缩等步骤估计载体图像的 DCT 系数直方图，根据待测图像直方图和估计直方图的相关改变量估计 F5 算法的嵌入量。该方法能有效检测低至 10%的嵌入量，但对具有特殊网格结构的图像无效。

Fridrich 等对待测图像进行 Outguess 嵌入操作，根据载体图像与隐藏图像像素块边界的增量差来估计嵌入算法 Outguess 的嵌入量。该方法无须确定阈值，对不可以由嵌入秘密消息的长度预见图像宏观改变量的情况以及以 DCT 系数的增/减量作嵌入算法的无效。

DCT 域隐写分析主要围绕 DCT 系数的统计特性及其对空域像素的影响进行研究，包括了对载体图像 DCT 系数的估计及空域像素块不连续性的计算。研究的方法经历了从简单的一阶统计分析到采用比较复杂的实验手段来获得相关变化量的过程，总体来说存在适用性较低、实用性不高等不足。

DWT 域隐写分析的研究报道较少，刘绍辉等针对 DWT 域 QIM 嵌入算法，提出了基于 DFT 域能量差分的检测算法，平均检测率达到 90%，这是检测 DWT 域隐写术的一个有益尝试。

2. 通用隐写分析方法

由于通用隐写分析的攻击对象包括了空域与变换域隐写术，是隐写分析领域不可忽视的部分。

Avcibas 等提出的 IQM's(Image Quality Metrics)方法，采用变量分析技术来分析和选取可用于区分载体图像和隐藏图像的质量度量，根据选取的图像质量特征采用多元回归的方法对图像进行分类。该方法能有效检测多种隐写方法，但是需要对分类器进行训练，性能一般。

Farid 等采用 QFM 分析图像小波域系数及其预测误差的高阶统计量，再分别采用 Fisher 线性判别式、线性与非线性支持向量机来判别和归类，对 DCT 域隐写术和以自然图像为载体的隐写术效果较好。但该方法需要对分类器进行训练，对嵌入量较低的空域隐写术和 OutGuess 的检测无效。

通用隐写分析主要围绕嵌入秘密消息前后待测图像的总体、局部、相关等特征值及具有训练模式的判别方法进行研究，但是通用特征的选取和阈值的确定非常困难，而且复杂度偏高、实用性不强、准确性较低，无法控制虚警率和漏报率。

3. 针对 JPEG 图像的隐写分析方法

根据特征选取与嵌入算法的关系，隐写分析方法可分为专用型和通用型两种。前者主要根据载密图像特征的改变，来提取专有特征进行检测，检测率较高，但实用性不强，只对特定的隐写术有效；后者主要是寻找独立于嵌入算法的统计特征向量，根据载体统计特性的变化判断是否含有隐密信息，它对一系列的隐写算法都有效，实用性较高，但整体检测率较弱。下面介绍针对 JPEG 图像的几种典型隐写分析算法。

1) 基于 χ^2 检测法

由于隐密信息的嵌入，载密图像和原始图像的 DCT 系数直方图的分布会发生改变，统计(卡方统计) 攻击就是通过观察测试图像的统计直方图来检测图像中是否含有隐密信息。该方法的关键点在于构造隐密信息的理论频率分布。对 DCT 系数来说，最低位为 1 的系数和最低位为 0 的系数构成一个系数对，成为 POV(Pairs of Value)，那么嵌入秘密信息前后最低位为 1 和 0 的这对系数之和不变，也就是这对 POV 的值保持不变。由于嵌入信息服从均匀分布，那么秘密信息嵌入前后 POV 的个数也保持不变。

以 h_{2i}表示 DCT 系数值为 $2i(i\neq0)$的数目，令 $n_i=h_{2i}$，经 Jsteg 算法处理后，n_i 期望值 n_i'为：

$$n_i'=\frac{h_{2i}+h_{2i+1}}{2} \tag{7-10}$$

则统计量

$$\chi^2(k-1)=\sum_{i=1}^{k}\frac{(n_i-n_i')^2}{n_i'} \tag{7-11}$$

的分布渐进$(k-1)$个自由度为 χ^2 分布。$n_i=n'$的概率为：

$$p = 1 - \frac{1}{2^{\frac{k-1}{2}}\Gamma\left(\frac{k-1}{2}\right)}\int_0^{\chi^2(k-1)} e^{-\frac{i}{2}} t^{\frac{k-1}{2}-1} dt \tag{7-12}$$

若概率 p 接近于 1，说明有隐密消息嵌入，若 p 值非常小，甚至接近于 0，则说明没有隐密消息嵌入。

在嵌入率较高的情况下，会更好地满足大数定理；此外，由于颜色对频度差与颜色对的频度比一般很小，密度函数曲线前面部分比较平坦，所以检测率也会较高。但该方法只适用于顺序信息的隐写，Westfeld 随后在 2002 年提出的通用卡方统计方法应用范围要更广一些。

2）空域 LSB 算法的 RS 分析算法

J. Fridrich 在 2001 年针对空域 LSB 置换算法提出了 RS 分析算法。RS 算法利用图像的空域相关性导出敏感对偶量，来检测图像中是否含有隐藏信息。

定义鉴别函数 f 来描述图像的空间相关性，函数 f 对一个像素组 $G=(x_1,x_2,\cdots,x_a)$ 指定一个实数 $f(x_1,x_2,\cdots,x_a)\in R$，函数定义如下：

$$f(x_1,x_2,\cdots,x_a) = \sum_{i=1}^{a-1} |x_{i+1} - x_i| \tag{7-13}$$

该函数用来度量像素组 G 内部的不平滑性：像素组 G 越不平滑，函数的值就越大，LSB 嵌入必然增大 G 的不平滑性。传统的 LSB 嵌入过程可以用置换函数 F_1 来描述：F_1：$0\leftrightarrow 1, 2\leftrightarrow 3,\cdots,254\leftrightarrow 255$ 改变灰度值 X 的 LSB 相当于对 x 利用置换函数 F_1。同时定义一个置换函数的对偶概念，移位 LSB 置换函数 F_{-1}：$-1\leftrightarrow 0, 1\leftrightarrow 2,\cdots,255\leftrightarrow 256$。$F_1$ 同 F_{-1} 关系为：

$$F_{-1} = F_1(x+1) - 1$$

同时定义 F_0 为自身置换函数，$F_0(x)=x$。

根据置换函数改变像素组鉴别值的方式定义了 3 种像素组：①常规类 R，如果 $f(F(G))>f(G)$；②异常类 S，如果 $f(F(G))<f(G)$；③不变类 U，如果 $f(F(G))=f(G)$。其中 $F(G)$ 代表对象组 $G=(x_1,x_2,\cdots,x_n)$ 中对每一个像素应用置换函数。通常对象组 G 中不同像素应用不同类型的置换函数，可以通过指定一个掩码算子 M，M 是元素值为 $-1,0,1$ 的 n 元组，因此定义：

$$F(G) = (F_{M1}(x_1), F_{M2}(x_2),\cdots,F_{Mn}(x_n)) \tag{7-14}$$

原始图像中，像素组 G 使用置换函数通常会使鉴别值增加，若把图像分为若干组，常规组将大于异常组。对非负掩码算子 M 来说，定义 R_M 常规类像素组个数与所有像素组的百分比，S_M 表示异常类像素组与所有像素组百分比，从而 $R_M+S_M\leqslant 1$ 和 $R_{-M}+S_{-M}\leqslant 1$。RS 分析算法的零消息假设是：对于载体图像，R_M 的值近似接近于 R_{-M}，S_M 近似等于 S_{-M}，即：

$$R_M \cong R_{-M}$$
$$S_M \cong S_{-M}$$

LSB 平面的随机变化使得 R_M 与 S_M 之间的差异随着嵌入长度的增加越来越小，改变 LSB 平面 50% 的像素之后，就有 $R_M\cong S_M$；但是 LSB 平面的随机变化对于 R_{-M} 和 S_{-M} 有相反的影响，它们之间的差异随着嵌入信息长度的增加而增大。

假设在图像的 LSB 平面嵌入长度为 p（像素的百分比）的秘密信息，Fridrich 经过大

量实验发现，R_{-M}和S_{-M}相对于P的变化曲线可以很好地利用直线进行建模，用二次多项式近似地逼近R_M与S_M相对P的变化曲线。因此可以通过计算R和S像素组的个数，基于各种约束关系，精确地计算P的值。

4. 针对 F5 的隐写分析算法

针对 F5 隐写术，Fridrich 提出了相应的隐写分析方法，它能够检测是否存在隐藏信息，并能估算隐藏信息长度。该算法分为两步：首先确定区分统计量T，T与被修改的 DCT 系数总数有关；然后确定统计量T的基值。

以$h_{kl}(d)$表示在载体图像所有 8×8 的 DCT 矩阵的(k,l)位置绝对值为d的 DCT 系数总和，$H_{kl}(d)$表示在载密图像相应位置绝对值为d的 DCT 系数的数目。若 F5 算法改动了n个 DCT 系数，则一个非 0 的 DCT 系数被改动的概率为：$\beta=n/p$，其中p为非 0 的 DCT 系数的总数。因为$F5$算法中系数的选择是随机的，所以$H_{kl}(d)$的期望值可表示为：

$$H_{kl}(d)=(1-\beta)h_{kl}(d)+\beta h_{kl}(d+1),\quad d>0 \tag{7-15}$$

$$H_{kl}(0)=h_{kl}(0)+\beta h_{kl}(1),d=0 \tag{7-16}$$

以$h'_{kl}(d)$表示对载体图像$h_{kl}(d)$的估计，利用最小均方误差估计可得β的最小值与$h'_{kl}(d)$和$H_{kl}(d)$的关系式：

$$\beta_{kl}=\arg\min\{[H_{kl}(0)-h'_{kl}(0)-\beta h'_{kl}(d)]^2+[H_{kl}(1)-(1-\beta)h'_{kl}(1)-\beta h'_{kl}(2)]^2\} \tag{7-17}$$

推导可得：

$$\beta_{kl}=\frac{h'_{kl}(1)[H_{kl}(0)-h'_{kl}(0)]+[H_{kl}(1)-h'_{kl}(1)]\cdot[h'_{kl}(2)-h'_{kl}(1)]}{h'^2_{kl}(1)+[h'_{kl}(2)-h'_{kl}(1)]^2} \tag{7-18}$$

最终β值为所选低频 DCT$(k,l)\in\{(1,2),(2,1),(2,2)\}$的平均值：

$$\beta=\frac{1}{3}(\beta_{12}+\beta_{21}+\beta_{22}) \tag{7-19}$$

该算法的关键是准确地估计出载体图像的$h'_{kl}(d)$。获得基准图像分为三步：

① 将隐写图像解压到空域；

② 利用 4 像素在横竖两个方向上对隐写图像进行裁减；

③ 利用与隐写图像相同的量化矩阵进行压缩。

5. 基于小波特征函数统计矩的隐写分析

基于小波特征函数统计矩的隐写分析法是一种通用型的隐写分析方法，它使用小波子带的特征函数统计矩作为隐写分析的特征。一个特征函数f_k的n阶统计矩被定义为：

$$M_n=\frac{\sum_{k=0}^{\frac{N}{2}}f_k^n\mid H(f_k)\mid}{\sum_{k=0}^{\frac{N}{2}}\mid H(f_k)\mid} \tag{7-20}$$

其中，$|H(f_k)|$是特征函数的幅值，即图像直方图的 DFT。根据离散形式的 Chebyshev

不等式，隐写技术嵌入隐秘信息后，M_n 的值将下降。

表 7-1　基于直方图的矩与基于特征函数的矩的比较

两种方式比较	矩的阶次		
	$n=1$	$n=2$	$n=3$
基于直方图的 n 阶矩：$\int_0^{\infty} \mid x \mid^n \{d(x)\}\mathrm{d}x$ 服从高斯分布的情况：$h(x)=\frac{1}{\sigma\sqrt{2}}\mathrm{e}^{\frac{x}{2\sigma^2}}$	$\frac{\sqrt{2}\sigma}{\sqrt{\pi}}$	σ^2	$\frac{2\sqrt{2}\sigma^3}{\sqrt{\pi}}$
基于特征函数的 n 阶矩：$\int_0^{\infty} \mid f \mid^n \{h(x)\}\mathrm{d}x$ 服从高斯分布的情况：$H(f)=\mathrm{e}^{\frac{\sigma^2 f^2}{2}}$	$2\left(\frac{1}{\sigma}\right)^2$	$\sqrt{2}\left(\frac{1}{\sigma}\right)^3$	$4\left(\frac{1}{\sigma}\right)^4$

在表 7-1 中，对直接根据图像直方图计算的前 3 阶矩与根据图像直方图特征函数计算的前 3 阶矩做了一个比较。具体地说，基于特征函数的 n 阶矩相当于 $\left(\frac{1}{\sigma}\right)^{n+1}$，而基于直方图的 n 阶矩相当于 σ^n。因此，基于特征函数的矩对隐密信息的标准方差 σ 的改变更为敏感。

思　考　题

7.1　什么是数字图像？数字图像取证有哪些特点？

7.2　数字图像内容安全与哪些学科之间有关联？它们之间如何相互联系和影响？

7.3　数字图像安全常见攻防手段有哪些？

7.4　简述数字图像加密技术常用方法。

7.5　查阅一个数字图像加密算法，写一篇阅读报告。

7.6　仿真完成一个数字图像水印算法。

7.7　数字图像水印技术今后的发展方向是什么？

7.8　检索最新的数字图像水印和数字图像隐写文献，写一篇阅读报告。

7.9　作为数字图像完整性取证的隐密分析取证技术分为几个阶段，目前所达到的技术主要集中在哪个阶段，要想取证结果作为法庭证据则需要取得怎样的结果？

7.10　比较空域隐写分析与变换域隐密分析的异同。

7.11　查阅 JPEG 文件格式，分析在 JPEG 文件格式中可能的隐密方法，并给出相应的隐密分析取证方法。

参 考 文 献

[1]　董刚，张春田. 信息安全领域的一种新技术——数字水印技术. 天津通信技术，2001，6(2)：34-38.

[2]　丁玮，齐东旭. 数字图像变换及信息隐藏与伪装技术. 计算机学报，1998，21(9)：838-843.

[3] 周瑞辉,荆继武.信息安全的新兴领域——信息隐藏.计算机应用研究,2001,7：6-8.
[4] 刘峰,张鹏.信息隐藏技术及其应用.天津通信技术,2001,5(1)：1-4.
[5] 祁明,刘迎风.信息隐藏与数字水印技术及其应用.通信技术,2001,6：16-19.
[6] 易开祥.数字图像加密与数字水印技术研究.浙江大学博士学位论文,2001.
[7] Cao Z F. A threshold key escrow based on public key cryptosystem. Science in China Seires E: Technological Sciences,2001,44(4):441-448.
[8] 冯登国,吴文玲. 分组密码的设计与分析.北京:清华大学出版社,2000.
[9] Protopopescu V A,Santoro R T,Tolliver J S. Fast and secure encryption-decryption method based on chaotic dynamics. U. S. Patent,Patent number：547951,1995.
[10] Short K M. Unmasking a modulated chaotic communications Scheme. International Journal of Bifurcation and Chaos,1996,6(02)：367-375.
[11] Yang T, Yang L B, Yang C M. Application of neural networks to unmasking chaotic secure communication. Physica D：Nonlinear Phenomena,1998,124 (1-3)：248-257.
[12] Ker A D. Improved detection of LSB steganography in grayscale images. In Proceedings of 6th Information Hiding Workshop,Lecture Notes in Computer Science,2004,97-115.
[13] 高婷婷.基于混沌的数字水印算法的研究.重庆大学硕士学位论文,2004.
[14] Li T Y,Youke J A. Period three implies chaos. The American Mathematical Monthly,1975,82 (10):985-992.
[15] Naor M,Shamir A. Visual cryptography. in Advances in Cryptology：EUROCRYPT' 94 (A. De Santis,ed.),of Lecture Notes in Computer Science,Springer. 1995,950:1-12.
[16] Simmons G J. Prisoners'problem and the subliminal channel. In Proceedings of CRYPT083 • Advances in cryptology. Santa Barbara,USA,1984,51-67.
[17] Provos N. Defending against statistical steganalysis. In 10th USENIX Security Symposium, Washington DC,USA,2001,323-336.
[18] Kawaguchi E,Eason R O. Principle and Application of BPCS-Steganography. In Proceedings of SPIE：Multimedia Systems and Applications,1998,3528:464-472.
[19] Wpham D. Jsteg. http://zooid. Org/-paul/crypto/JSteg.
[20] Latham A. Jphide. http://linux01. gwdg. de/-alatham/stego. html.
[21] Westfield A. F5—A Steganographic Algrithm High Capacity Despite Better Steganalysis. IH2001,2001,289-302.
[22] Lu P,Luo X. Tang X Q,et al. An improved sample pairs method for detection of LSB embedding. In Proceedings of the 6th Information Hiding Workshop,Lecture Notes in Computer Science, 2004,3200：116-127.
[23] Chen B,Womell G W. Quantization Index Modulation Methods for Digital Watermarking and Information Embedding of Multimedia. Journal of VLSI Signal Processing,2001,27(1-2)：7-33.
[24] Cox I J,Kilian J,Leighton T,et al. Secure spread spectrum watermarking for multimedia. IEEE Transactions on Image Processing,1997,6(12):1673-1678.
[25] Marvel L,Boncelet C,Retter C. Spread Spectrum Image Steganography. IEEE Transactions on image processing,1999,8(8)：1075-1083.
[26] Chandramouli R. A mathematical framework for active steganalysis. Multimedia Systems,2003, 9(3),303-311.
[27] Trivedi S,Chandramouli R. Active steganalysis of sequential steganography. Proceedings of

SPIE on Security and Watermarking of Multimedia Contents V,2003,5020:123-130.
[28] Trivedi S, Chandramouli R. Secret key estimation in sequential steganography. IEEE Transactions on Signal Processing,2005,53(2):746-757.
[29] Westfeld A, Pfitzmann A. Attacks on steganographic systems. Lecture Notes in Computer Science,1999,1768: 61-76.
[30] Provos N. Defending against statistical steganalysis. In: Proceedings of the 10th USENIX Security Symposium, Washington, D. C. , USA, Aug. 13-17,2001,323-335.
[31] Fridrich J, Goljan M, Du R. Attacking the OutGuess. In: Proceedings of the ACM Workshop on Multimedia and Security, Juan-les-Pins, France, Dec 2002,3-6.
[32] Fridrich J, Goljan M, Du R. Detecting LSB steganography in color and gray-scale Images. IEEE Multimedia,2001,8(4):22-28.
[33] Westfeld A. F5—a steganographic algorithm. Lecture Notes in Computer Science,2001,2137: 289-302.
[34] Fridrich A, Goljan M, Hogea D. Steganalysis of JPEG image: breaking the F5 algorithm. In Proceedings of 5th International Workshop on Information Hiding, Noordwijkerhout, Netherlands,2002,310-323.
[35] Kawaguchi K, Eason R O. Principle and application of BPCS steganography. In Proceedings of SPIE on Multimedia Systems and Applications, Boston 1998,3528:464-472.
[36] Zhang X, Wang S. Statistical analysis against spatial BPCS steganography. Journal of Computer Aided Design and Computer Graphics,2005,17(7):1625-1629.
[37] Wu M, Ho Y, Lee J. An iterative method of palette-based image steganography. Pattern Recognition Letters,2004,25(3): 301-309.
[38] Fridrich J, Rui D. Secure steganographic methods for palette images. The 3rd Information Hiding Workshop, Lecture Notes in Computer Science,2000,1768:47-66.
[39] Ker A D. Resampling and the detection of LSB matching in colour bitmaps. Proceedings of the SPIE on Security, Steganography, and Watermarking of Multimedia Contents Ⅶ, 2005, 5681: 1-15.
[40] Sharp T. An implementation of key-based digital signal steganography. In: Proceedings of Information Hiding Workshop, Lecture Notes in Computer Science,2001,2137:13-26.
[41] Harmsen J, Pearlman W. Steganalysis of additive-noise modelable information hiding. In Proceedings of SPIE on Security Watermarking Multimedia Contents,2003,5020: 131-142.
[42] Ker A D. Steganalysis of LSB matching in grayscale images. IEEE Signal Processing Letters, 2005,12(6): 441-444.
[43] Fridrich J, Lisoněk P. Grid coloring in steganography. IEEE Transactions on Information Theory,2007,53(4):1547-1549.
[44] 许郡. 图像数字水印技术研究. 扬州大学硕士学位论文,2009.
[45] 刘鸿霞,夏春和. 图像隐写分析现状研究. 计算机工程与设计,2006,1: 21-25.
[46] 陈铭,张茹,钮心忻,杨义先. 隐写分析技术研究综述. 计算机应用,2008,28: 21-33.
[47] 高婷婷. 基于混沌的数字水印算法的研究. 重大大学硕士学位论文,2004.

chapter 8

第8章 数字音频内容安全

本章学习目标

数字音乐的发展使得数字音频的安全成为当前数字内容安全领域的一个重要研究内容。本章将对数字音频内容安全的有关概念和方法进行介绍，主要包括数字音频内容加密、数字音频隐写、数字音频隐写分析以及数字音频取证等方面。

通过本章学习，应掌握以下内容：

(1) 数字音频内容加密技术。

(2) 数字音频水印技术。

(3) 数字音频隐写分析技术。

(4) 数字音频取证技术。

8.1 数字音频内容安全基本概念

截止到2009年底，我国互联网使用人数已达3.84亿。而在所有网络应用中，使用频率最高的网络应用是网络音乐，使用率高达83.5%。这意味着，仅在我国拥有收听、下载、分享网络音乐习惯的用户数就高达3.2亿。由于数字媒体产品具有容易复制、保存、篡改和传播的特点，非法上传、下载、篡改网络音乐时有发生。由于没有从技术上彻底解决如何防止信道窃听和实施盗版源头追踪这两个问题，盗版现象仍然大量存在。如何保证数字音频的内容安全成了一个重要的难题。

要研究数字音频的安全问题，就需要了解数字音频区别于其他信息载体的特点，需要了解人类听觉的特殊性。因此了解语音的基础知识，以及经常使用的语音处理方法，对于设计好的音频保护方案有着重要的指导意义。如果能够很好地利用语音的各种处理方法，可以设计出对各种处理方法具有鲁棒性的音频数字水印和隐写算法。

8.1.1 音频信号的数字表示

自然界中的音频信号是幅度随时间而变的一维连续信号，不仅在时间上是连续的，而且在幅度上也是连续的。它的频率范围一般是20～20000Hz，称为模拟信号，计算机是无法对这种模拟信号进行处理的。

计算机只能处理时间上和幅度上都是有限的信号，也就是数字信号，要对声音信号进行计算机处理，就必须对声音信号进行数字化（即 A/D 转换）。

数字化实际上就是采样和量化、编码。连续时间的离散化通过采样来实现，在某些特定的时刻对模拟信号进行测量叫做采样（sampling），每隔相等间隔采样一次，这种采样称为均匀采样，下面使用的声音文件都是均匀采样。连续幅度的离散化通过量化来实现，也就是把信号的强度划分成不同等级。如果幅度的划分是等间隔的，就称为线性量化，否则就称为非线性量化。采样的精度、样本的大小是用每个声音样本的比特率（b/s）表示的，它反映度量声音波形幅度的精度。脉冲编码调制（Pulse Code Modulation，PCM）是最简单的波形编码方式。

8.1.2 音频文件的存储格式

要在计算机内播放或处理音频文件，需要对声音文件进行数、模转换，这个过程同样由采样和量化构成。人耳所能听到的声音，最低的频率是从 20Hz 起一直到最高频率 20kHz，而 20kHz 以上的声音人耳是听不到的，因此音频的最大带宽是 20kHz，所以采样速率需要介于 40kHz 与 50kHz 之间，而且对每个样本需要更多的量化比特数。音频数字化的标准是每个样本 16 位（即 96dB）的信噪比，采用线性脉冲编码调制 PCM，每一量化步长都具有相等的长度。在音频文件的制作中，正是采用了这一标准。

在互联网和各种机器上，声音文件的格式很多，比较流行的有以 WAV、AU、AIFR、SND 为扩展名的文件格式。其中 WAV 格式主要用在 PC 上，AU 格式多用于 UNIX 工作站，AIFR 和 SND 则主要用于苹果机和 SGI 工作站，而在互联网上绝大多数是 MP3 格式的文件。MP3 格式是 MPEG-1 标准的第三层，因其具有非常高的压缩率而得到广泛应用。但由于 MP3 文件是熵编码，无法直接得到音频信号的原始幅值，也就无法直接对其作信号处理。因此使用的水印载体是 Windows NT 系统使用的 WAV 格式音频文件。

对于现在网络上流行的 MP3 文件，若要对其进行版权保护，可以将 MP3 文件转化为 WAV 格式文件，将水印信息嵌入到 WAV 格式文件，再将 WAV 格式文件转化为 MP3 文件。这要求水印对这一转化过程是鲁棒的，因为从 WAV 格式文件转化为 MP3 文件是一个有损压缩的过程。此外，采样频率会影响数据隐藏，因此它给出了可用频谱的上限（如果信号的采样频率为 8kHz，则引入的修改分量的频率不会超过 4kHz）。对于大多数已有的数据隐藏技术而言，可用的数据空间与采样频率的增长至少呈线性关系。

8.1.3 音频信号的传输环境

在实际使用中，含水印的音频信号从编码到解码之间有多种可能的传播途径。下面列举最普通的四种情形。

（1）声音文件从一台机器通过存储介质或网络复制到另一台机器，其中没有任何形式的改变。编码方和解码方的采样率完全一样。

（2）信号仍然保持数字的形式，但采样率发生变化。这一变化保持了大多数信号的幅度和相位值，但是改变了信号的时域特性。

(3) 信号被转换为模拟形式,通过模拟线路进行传播,然后在终端被重新采样。在此过程中信号的幅度、量化方式和时域采样率都得不到保持,但在这种情形下信号的相位值可以得到保持。

(4) 信号在空气中传播,经过麦克风重新采样。这时的信号受到未知的非线性改变,会导致相位改变、幅度改变、不同频率成分的漂移和产生回声等。

通过比较四种情形,可以知道在第一种情形下,音频信号在传输中没有改变。而在第四种情况下音频信号在传输中发生的变化最大。

8.1.4 人类听觉特性

人类听觉系统对音频文件中附加的随机噪声敏感,并能觉察出微小的扰动。人耳听觉系统具有复杂的特性,涉及有关心理声学和生理声学方面的问题,通常需用非线性模型表示。心理声学的一个重要特性就是人耳的掩蔽效应,声音信号在人的听觉系统中会经过非线性加工,掩蔽效应正是由于这种听觉的非线性引起的常见心理声学现象。

首先,人的听觉具有掩蔽效应。掩蔽效应是指一个较弱但可以听到的声音由于另外一个较强的声音的出现而变得无法听到的现象。掩蔽的效果依赖于掩蔽音和被掩蔽音的时域和频域特性。因此听觉掩蔽可以分为频域掩蔽和时域掩蔽。频域掩蔽指在频域发生的掩蔽现象。如果在一定频率范围内,同时存在能量相差一定程度的一强一弱两个信号时,弱音不被人耳察觉,即被强音"掩蔽"掉,则较强的音称为掩蔽音,弱音称为被掩蔽音。把一个纯音调作为目标,如果它的声压级低于绝对阈值(安静时的听觉阈值),它是听不见的。由于一个较强信号的存在,听觉阈值不同于安静时的阈值,在接近较强信号频率的频率处,听觉阈值被提高,新阈值称为掩蔽阈值,当信号的声压级低于掩蔽阈值时,它被掩蔽。一个掩蔽音的掩蔽阈值依赖于频率、声压级,以及掩蔽和被掩蔽信号的纯音或噪音特性。用一个宽带的噪音掩蔽一个纯音比用一个纯音掩蔽一个宽带的噪音要容易。而且,信号频率愈高就愈易被掩蔽。时域掩蔽包括向前掩蔽和向后掩蔽。向前掩蔽是指较强的掩蔽音出现之前较弱的被掩蔽音无法听到,向后掩蔽是指较强的掩蔽音消失后较弱的被掩蔽音无法听到。一般,向前掩蔽发生在掩蔽音出现前 5～20ms,向后掩蔽发生在掩蔽音消失后 50～200ms。频域和时域掩蔽效应有各自的特性及局限,频域掩蔽效应局限在频率域,而时域掩蔽效应则局限在时间域。

其次,人耳对声音信号的绝对相位不敏感,而只对其相对相位敏感。另外,人耳对不同频段声音的敏感程度不同,通常人耳可以听到 20Hz～18kHz 的信号,其中对 300～3400Hz 范围内的信号最为敏感,幅度很低的信号也能被听见,而在低频区和高频区,能被人耳听见的信号幅度要高得多。即使对同样声压级的声音,人耳实际感觉到的音量也是随频率而变化的。要提高水印信号的不可感知性,可将水印信号或私密信息加载在掩蔽阈值较高的高频部分。

8.2 数字音频内容加密技术

8.2.1 数字音频加密技术简介

音频加密技术包括两个元素：算法和密钥。算法是将数字音频信息与一串数字（密钥）的结合，产生不可理解的密文的步骤，密钥是用来对数据进行编码和解码的一种算法。在安全保密中，可通过适当的密钥加密技术和管理机制来保证网络的信息安全。由于密码体制包括对称密钥体制和非对称密钥体制两种，相应地，音频数据加密技术也分为两类，即音频对称加密（私人密钥加密）和非对称加密（公开密钥加密）。

8.2.2 数字音频加密技术分类

按加密算法在数字音频编码压缩过程的所处位置，音频数据加密方法可分为完全加密（位置①和位置③）和选择性加密（位置②），见图 8-1。

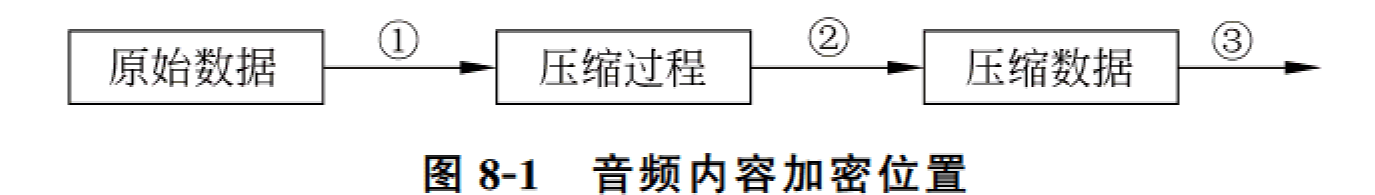


图 8-1 音频内容加密位置

1. 完全加密

具体来说，完全加密又可分为对原始多媒体数据（即压缩前）的加密和压缩后数据的加密。该类加密方法把多媒体数据当成普通二进制数据，直接使用传统的 RSA、DES 等加密算法。

音频的完全加密不需要考虑音频的编码格式，将全部的数据加密，由于加密的数据量大，故具有较高的安全性。但是当音频数据量较大时，这种方法的效率较低，同时加密后的音频因为数据格式的改变而不能支持直接操作。

2. 选择性加密

选择性加密（selective encryption）又称部分加密（partial encryption）。该类算法通过对精心选择的部分重要数据进行加密实现所需的加密效果。部分加密技术最初见于 1995 年，设计的初衷主要是为了解决视频点播（Video On Demand，VOD）和 MPEG-1 的加密传输问题。通过结合多媒体编解码过程，挑选出一些对多媒体解码影响大、带有丰富信息的参数，然后对参数进行加密。后来选择性加密技术逐渐应用到图像和音频处理。选择性加密算法大大降低了加密的数据量，提高了加密速度，但降低了算法的安全性。如仅仅置乱 DCT 块内系数无法抵抗已知明文攻击。该类算法适用于实时性要求较高、但安全要求不太严格的应用场合。

另外，选择性加密可以控制加密程度，能应用于“预览版”、“体验版”多媒体产品的编辑。常见方法有只加密音频的某个区域、轻微加密音频的某一部分、模糊化部分原始音

频等。相比于通过图像背景叠加、加背景噪声、调视频亮度等常见的多媒体编辑方式来达到多媒体"失真"的方法，选择性加密技术具备简单、可逆性、方便快捷、原理通用和更安全的特点。要注意的是，加密后的多媒体数据必须符合文件的格式标准，即加密后的多媒体依旧能被标准的解码设备所读取。

8.3 数字音频隐写与水印技术

音频隐写术和数字水印的研究是随着信息技术的发展受到重视并蓬勃发展起来的。音频信息在网络上的传递、发布和扩散带来了一系列的问题和应用需求。从总体上来说可以分为两大部分：伪装式保密音频信息传递（即隐写术）和音频信息的版权保护（即数字水印）。隐写术主要应用在需要安全保密通信的部门，利用音频信息中的冗余空间携带隐蔽信息，达到秘密信息伪装传递的目的；同时，隐写术还要研究其对立面——隐秘信息的分析和检测，这与密码编码学和密码分析学的关系是类似的，信息隐藏与隐秘信息分析是一对矛盾的统一体，它们相互对立又相互促进。数字水印从实质上说也是一类信息隐藏，但是其目的不是为了保密通信，而是为了标明载体本身的一些信息，如音频信息的创作者、版权信息、使用权限等一系列需要标明的信息，利用数字水印，还可以跟踪音频产品的非法传播和扩散，打击盗版。数字水印技术目前正处于持续深入发展的阶段，应用领域也在快速扩展。

8.3.1 音频数据中的常用隐写算法

在音频隐写术算法的研究上，许多方法都是直接从图像隐写术中借鉴过来的，当然也有一些是音频特有的。1996 年，W. Bender 等在 IBM 的杂志上发表的"数据隐藏技术"一文中，较为系统地对音频为载体的低位嵌入法（LSB）、相位嵌入法、直接序列扩展频谱嵌入法（Direct Sequence Spread Spectrum，DSSS）和回声嵌入法作了介绍。

近年来有关音频信息隐藏技术方面的研究发展很快，很多基于 HAS 的方法被提出，常用的方法有以下几种。

1. 最不重要位法（LSB）

LSB 是信息隐藏的最简单、有效的一种方法，通过将原始数据的部分样本值的最低比特位或最低几个比特位用代表秘密数据的二进制位替换达到将秘密信息隐藏到音频中的目的。在接收端，只需要从相应位置提取出秘密信息比特即可。

LSB 算法简单易实现，信息嵌入和提取的速度快，可以隐藏的数据量大，但是其安全性很差，攻击者只需要对信道简单地加上噪声干扰或者对数据进行亚采样和压缩编码等处理都会造成整个隐秘信息的丢失。为了加大检测秘密数据的难度，可以用一段伪随机序列来控制嵌入秘密的位置，并采用不同的加密方式对数据本身和嵌入过程进行加密。为了提高鲁棒性同时保证隐蔽性，可在嵌入过程中根据音频信号的能量进行数据嵌入位的选择，同时确保原信号的最小嵌入失真。

2. 相位隐藏法

相位编码是利用人类听觉系统对声音绝对相位的不敏感,但对相对相位敏感的特性进行信息隐藏。

在相位编码中,隐藏的数据是用相位谱中特定的相位或相位变化来表示,可将音频信号分段,每段做DFT,数据只隐藏在第一段中,用代表秘密信息的参考相位替换第一段的绝对相位,为保证信号间的相对相位不变,所有随后信号的绝对相位也同时改变。接受端只要提取第一段的相位谱信息即可。系统参数包括段的大小和相位的变化量。为提高提取信息的准确性,一般取相位偏移值为$\pm\pi/2$。通常来说,相位隐藏方法的信道容量为8～32b/s。算法对载体信号的再采样具有鲁棒性,但对绝大多数音频压缩算法敏感。

3. 回声隐藏法

主要是利用了人耳听觉系统的另一个特性,即音频信号在时域的向后屏蔽作用,在离散信号中引入回声,通过修改信号和回声之间的延迟来编码水印信息,提取时,计算每个信号片断中信号倒谱的自相关函数,在时延上会出现峰值,对滤波、重采样、有损压缩等不敏感,但容易被第三方用回声检测的方法检测出来。

回声算法具有较好的透明性,但没有达到令人满意的正确提取率。为此,有人提出在信息提取时使用指数序列加权隐写数据段的改进方案,还提出了基于衰减系数的回声隐藏算法。

4. 变换域法

变换域法在图像水印中已经得到了广泛的应用,现在也越来越多地应用于音频水印中。这一方法的基本思想是通过将秘密信息嵌入到数字作品的某个变换域中达到将秘密信息嵌入到隐秘载体中最重要部分的目的,这样,只要攻击者不过分破坏隐蔽文件的可听懂度,嵌入信号中的隐秘信息就不会被删除。比较常见的变换域法有离散傅里叶变换法、离散余弦变换法、小波变换法、倒频谱域等。这些方法将秘密信息嵌入到频域变换的系数当中,并借鉴扩展频谱等技术对待隐藏信息进行有效的编码,从而提高了透明性和鲁棒性,同时还适当利用滤波技术消除信息隐藏可能引入的高频噪声,从而增加对低频滤波攻击的抵抗力。

基于变换域的方法用于音频信息隐藏可以更好地抵抗各种信号处理,而且还保持了对人类听觉的不可觉察性。

8.3.2 音频隐写工具

近年来,人们已经提出了不少成功的隐写方法。一些算法被陆续开发成隐写工具,其中不少可以免费获得,如在 http://jjtc.com/Steganography/toolmatrix.htm 上给出了数以百计的隐写软件。其中S-Tools、Hide4PGP等可以实现WAV声音信号的信息隐藏。考虑到信息传送的安全性,这些软件一般先对秘密信息进行数据压缩、加密预处理,

再将秘密信息嵌入到载体信号的最低位来达到信息隐藏的目的，隐藏容量由载体的长度决定。

下面以 S-Tools 为例，介绍可以实现音频和语音信息隐藏的软件工具。

S-Tools 采用了加密机制将信息隐藏在 BMP 或 GIF 图像文件中，也可以隐藏在 WAV 的声音文件中。现阶段支持的加密方法包括 IDEA、DES、3DES 等。

通常，Windows 中的 WAV 文件声音样本的位数是 8 位或者 16 位。当用 8 位来表示时，声音样本可能取值为 0～255 之间的整数，而对于 16 位，声音样本的取值是在 0～65535 的范围之内。

S-Tools 采用的是在声音信号的最不重要位隐藏文件。其基本算法原理如下：

假设声音样本信号包括以下字节：

132　134　137　141　121　101　74　38

则二进制值表示为：

10000100　10000110　10001001　10001101

01111001　01100101　01001010　00100110

要隐藏的二进制序列为 11010101(213)，则声音信号中的每个字节的 LSB 都会被这个字符的比特位代替，结果为：

133　135　136　141　120　101　74　39

二进制表示为：

10000101　10000111　10001000　10001101

01111000　01100101　01001010　00100111

可以看出，新的字节串和原始的字节串之间差别不大，人耳一般不能察觉出其中的不同，以上就是 S-Tools 的嵌入原理。S-Tools 是离散嵌入方案，它利用密钥随机选择信息隐藏的位置。

8.3.3 音频数字水印基本原理

数字水印模型是数字水印算法的基础，数字音频水印中常用的算法模型与图像水印等类似。图 8-2 是针对音频水印的算法模型，这个模型由 Voloshynovskiy 针对图像水印提出的，但是同样适用于音频水印。其中，水印编码环节负责隐秘信息加密，纠错编码。心理声学模型提供掩蔽阈值信息来确定水印最大可能嵌入强度。水印提取环节与水印嵌入环节相对应。水印检测环节负责判断水印的存在性，但是不提供水印的内容。水印

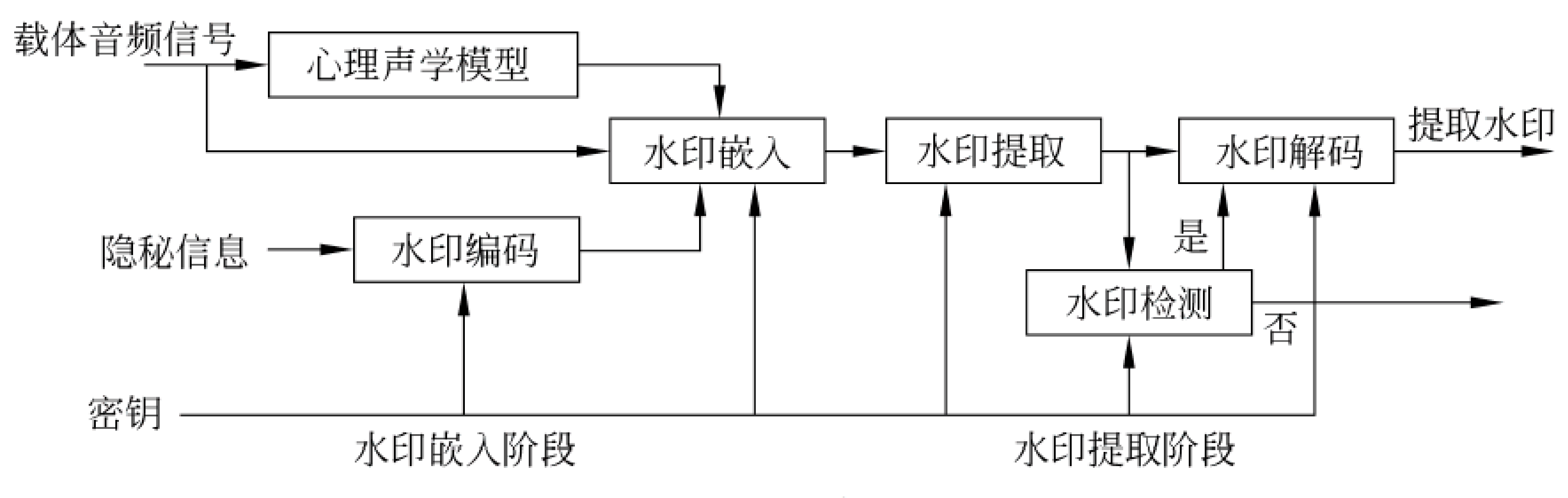


图 8-2　音频水印算法基本模型

解码环节负责提取隐秘信息。

这个模型是个基本模型，具体水印算法可能只包含其中的几个环节，例如回声隐藏模型没有明显包含心理声学模型环节，虽然它们间接利用了听觉系统感知模型特性。另外，如果水印信息本身是个伪随机序列，则不会包含水印解码环节。

8.3.4 数字音频水印的基本要求

要想成功地在数字音频媒体中嵌入水印，除了满足水印框架(GWF)的基本要求外，考虑到音频数据和人耳听觉的特性，还应注意以下几方面的要求：

(1) 对数据变换处理操作的鲁棒性。这就要求水印本身应能经受得住各种有意或无意的变换(攻击)。典型的变换有叠加噪声、数据压缩、滤波、重采样、几何变换、统计攻击等。

(2) 知觉相似性。数字水印是在对象中嵌入一定量的隐蔽信息，为使得第三方不易察觉这种嵌入信息，需谨慎选择嵌入方法使嵌入信息前后不产生可感知的变化。这种知觉相似性在理论上可用“知觉相似性函数”Sim(X,Y)来描述。数字音频中两个信号的相关性函数可用作相似性函数。

(3) 是否需要原始数据进行信息提取，这一要求将影响方案的用途和性能。根据数据嵌入和提取方案的不同设计，有些方案可以不需要借助于原始数据进行信息提取。

(4) 提取误码率。低误码率也是音频水印方案中的一个重要技术指标。因为一方面存在来自物理空间的干扰，另一方面信道中传输的信号会发生衰减和畸变，再加上人为的数据变换和攻击。

(5) 嵌入数据量指标。根据用途的不同，在有些应用场合中须保证一定的嵌入数据量。

8.3.5 数字音频水印的算法分类

音频水印的分类方法有不同的标准，从大的方面来分，可分为鲁棒型水印和脆弱型水印。

鲁棒型水印主要用于数字产品的版权保护，它必须保证对原始版权的准确无误的标识。因为数字水印时刻面临着用户或侵权者无意或恶意的破坏，因此，鲁棒型水印技术必须保证在宿主信号可能发生的各种失真变换的情况下，以及各种恶意攻击的情况下都具备很高的抵抗能力。

脆弱型数字水印主要用于数据的真伪鉴别和完整性鉴定，又称为认证。该水印技术在原始真实信号中嵌入某种标记信息，通过鉴别这些标记信息的改动情况，达到对原始数据完整性检测的目的。

如果从信号处理的角度分类，音频水印算法又可分为时间域算法、频域算法以及压缩域算法。

1. 时间域算法

时间域算法在时间域上将水印直接隐藏在数字音频信号。时间域水印算法的关键

是水印嵌入的位置，为了使嵌入的水印有更好的稳健性，水印的嵌入位置要充分利用人类听觉系统的特点。

典型的时间域嵌入方法是最低比特位 LSB 方法，通过把每个采样点的最低比特位用一个水印比特来代替，可以把大量的数据植入到音频信号中。这种方法的主要缺点是鲁棒性较差，如果不采用冗余技术，则水印信息很容易被噪声、重采样等所破坏，实用价值不大。有学者提出了基于音频段能量量化的时间域水印方案，水印提取无须原始音频参与。该算法的主要思想是根据即将嵌入的水印位对音频段的能量进行量化调制，以一定的比例修改音频段各采样点的幅值使此段能量变化为量化后的值。也可将音频信号划分为若干个包含相同采样点的段，每一段划分为若干个包含相同采样点的节，对每段前二节的能量进行比较，结合水印信号及 HAS 的掩蔽特性，采取不改变和缩小音频信号能量的方法，在数字音频中嵌入水印。

2. 频域算法

频域算法将对某一帧信号频域系数的修改扩散到该帧所有的时域采样点，而且，如果水印的频域嵌入只影响频域系数的幅度，检测/提取水印时可不要求水印信号的精确同步。常用频率域方法有 DFT、DCT、小波变换和 KLT 等。

基于 FFT 的水印加入技术通过对原始音频信号分段进行快速傅里叶变换，量化幅度加入水印；水印的提取则通过量化后的幅度所属的区间来判断。如将二维数字水印（灰度图像）编码成一维二进制序列并进行随机置乱，再对数字音频信号进行分段处理并依据人类听觉系统（HAS）择段作离散余弦变换（DCT），最后在 DCT 域内通过修改中高频 DCT 系数完成水印信息的自适应嵌入。

对基于特征点（信号短时平均幅度从低向高改变程度最大的点）检测的水印算法，主要利用了特征点的检测进行水印嵌入点的准确定位，应用离散余弦变换进行水印的嵌入。

采用基于离散余弦变换及奇异值分解的数字音频水印算法时，首先对二值水印图像进行奇异值分解求出奇异值，求出对角矩阵 $\boldsymbol{S}$，取其对角元素值，然后对音频信号进行离散余弦变换，将其分段并求出水印嵌入点，再对 $\boldsymbol{S}$ 的对角元素进行基于音频信号频率性质的调制处理，将经过调制的水印信号嵌入到音频信号变换域系数的幅值上。

另外，还可以利用离散小波变换的多分辨率特性，在小波变换的细节部分搜索局部极值点，通过修改局部极值及其相邻点的幅度值，实现水印的嵌入。

3. 压缩域算法

数字音频压缩技术的成熟，使得以 MP3 为代表的压缩格式的网络音乐得以在互联网上广泛传播。通常有三种方法可以得到带水印的压缩音频，如图 8-3 所示。

（1）在非压缩域进行，即先向非压缩原始音频中加入水印然后再压缩。通过采纳更为稳健的同步信号及其全新嵌入策略，提高音频水印的抗攻击能力；再结合听觉掩蔽特性自适应地确定量化步长，提高数字水印的不可感知性；特别地，这种方法对于最为普通的 MP3 压缩攻击具有极强的抵抗能力（尤其是高压缩比下）。

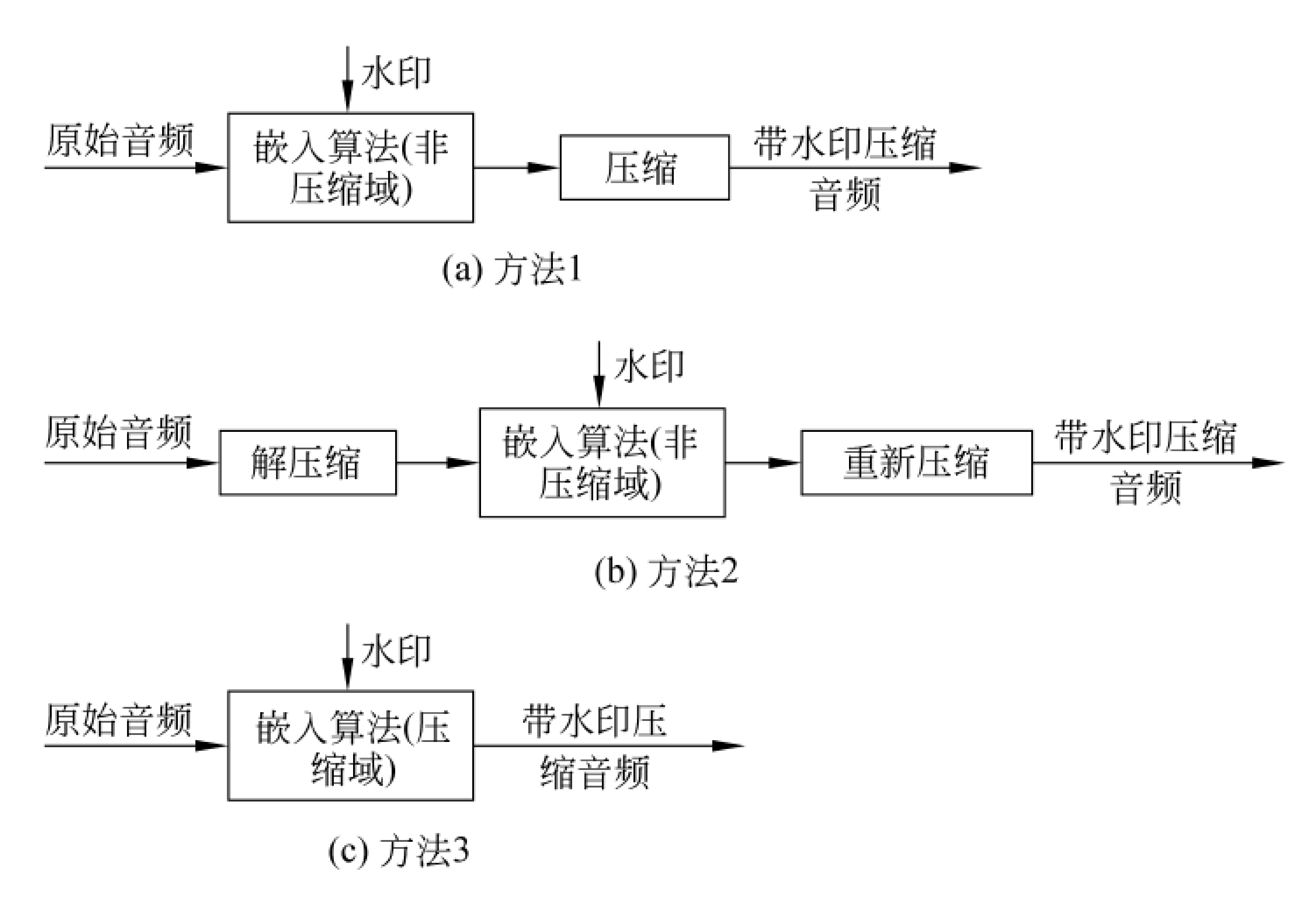


图 8-3 压缩域音频水印

(2) 首先将压缩格式的音频解压,然后将水印植入到非压缩域,最后带水印的音频内容再被重新压缩成带水印的压缩格式音频。例如,基于压缩音频内容的比特流水印算法,该算法首先根据编码算法将压缩音乐分段成音频帧,并解码到非压缩域,接着对每一帧进行特征提取和心理模型计算,根据提取的特征和计算出的掩蔽阈值,设计一个滤波器组来选择适合水印嵌入的候选帧,然后使用自适应多比特位跳跃将水印信息嵌入到选出的音频帧,再将嵌入水印后的音频帧重新编码压缩,最后将重编码后的音频帧和没有嵌入水印的音频帧重构生成带水印的压缩音乐。该方法可以提高水印的鲁棒性,但时间开销太大,因为压缩过程要花费很长时间,所以不适合在线交易和分发。

(3) 在压缩域上进行,水印直接加到 MPEG 音频比特流上,这使水印嵌入非常迅速,但鲁棒性较差,它同时又是真正意义上的压缩域水印方法,因为前两种本质上还是在非压缩域上进行水印嵌入的。如可以选择 MPEG 音频流的比例因子(scale factors)和 MPEG 编码的样本数据作为水印嵌入位置,将两种水印直接嵌到 MPEG 音频流中。

此外还有基于生理模型算法、基于音频内容算法、基于神经网络算法以及基于音频压缩标准与音频文件格式等算法。

8.3.6 常见数字音频水印算法

1. 音频水印时域算法

音频水印时域算法较少,其中最为主要的是 Basia、Pitas 提出的 LSB 方法和 W. Bender 提出的回声隐藏(Echo Hiding)算法。

1) LSB 算法

LSB 算法由 Basia、Pitas 等于 1996 提出。LSB 算法是一种最简单的水印算法,其主要方法是对音频信号进行采样,将采样值最不重要的位(通常为最低位)用代表水印的二

进制位代替，以达到在音频信号中嵌入水印数据的目的。

LSB算法的主要特点是嵌入及提取水印速度快，算法简单，容易实现，音频信号中可编码的数据量大；其缺陷是稳健性差。

2）基于回声的水印算法

回声(echo hiding)算法是一种经典算法，最初由W. Bender等于1996年提出。其主要方法是通过引入回声来将秘密数据嵌入到载体数据中，利用HAS的滞后掩蔽特性，即弱信号在强信号消失之后变得无法听见，它可以在强信号消失50～200ms作用而不被人耳觉察。

在回声隐藏的算法中，编码器将载体数据延迟一定的时间并叠加到原始的载体数据上以产生回声。编码器可以用两个不同的延迟时间来嵌入“0”和“1”。在实际的操作中，用代表0或1的回声内核与载体信号进行卷积来达到添加回声的效果。要想使嵌入后的隐秘数据不被怀疑，并且能使接收方以较高的正确率提取数据，关键在于回声内核的选取。每个回声内核具有四个可调整的参数：原始幅值、衰减率、1偏移量和0偏移量。

回声算法的特点是透明性好，可盲水印检测；但提取水印的正确率不能令人满意。

2. 音频水印变换域算法

1）相位水印算法

Bender W等于1996年提出的音频相位编码(phase coding)充分利用了人类听觉系统HAS的特性，即人耳对绝对相位不敏感性及对相对相位的敏感性。根据这一特性，用代表秘密数据位的参考相位替换原始音频段的绝对相位，并对其他的音频段进行调整，以保持各段之间的相对相位不变。

相位水印算法的特点是：当代表水印数据的参考相位急剧变化时，会出现明显的相位离差，会影响水印的隐蔽性以及增加水印解码的难度。当音频信号是较安静的环境时，嵌入的数据量较少。

2）离散余弦变换(DCT)算法

Wang Ye等于1998年提出了一种基于音频DCT变换域的水印算法，主要方法是：首先根据伪随机序列重新排列音频采样信号，对序列进行修正离散余弦变换(Modified Discrete Cosine Transform，MDCT)，通过对MDCT的系数进行改变以便嵌入水印，然后进行逆变换得到嵌入水印后的音频序列。

DCT算法的主要特点是选择变换系数(低频、中频或高频)，局部修改某些变换系数，以实现水印的嵌入。其透明性较好，能平滑功率谱密度，稳健性随所选频域嵌入系数而有所不同。

3）离散小波变换(DWT)算法

钮心忻等于2000年提出了一种利用小波变换的音频水印算法。该方法利用Daubechies-4小波基对原始语音信号进行L级小波分解，对L级的粗糙分量不予处理，对L级的精细分量进行处理，以嵌入水印。

陈琦等于2002年提出了利用小波变换将一枚签章的数字图像作为水印，嵌入到小波变换的第三层的精细分量中，并在信号嵌入时使用了检测同步信号，但检测时需要通

过原始音频信号进行比较才能获得水印。其主要特点是有较好的透明性和较强的健壮性。

3. 其他类型的音频水印算法

1）比特流水印算法

比特流水印算法由 Neubauer C. 等于 2000 年提出。水印系统完全工作在比特流域，输入和输出信号都是经过压缩编码的音频信号。其特点是复杂度较低、计算效率高、合成声音质量较好、稳健性较好。

2）压缩水印算法

Siebenhaar 等于 2001 年提出一种压缩水印方案，输出的是嵌入水印的音频比特流。特点是压缩和水印参数之间可实现最佳匹配，音频压缩和水印嵌入可同时处理，计算复杂度较低。

3）扰动调制水印算法

扰动调制水印算法由 Chen 和 Wornell 于 1999 年提出，并被应用于图像中。马田等对扰动调制进行了改进，将其应用于音频水印技术中。其基本思想是通过嵌入信息来调制量化器，对音频信号加上随机振动信号，然后再进行线性量化。其特点是由于采用了频域嵌入和多维量化器，在不完全同步的情况下，也能够在较大的容限内完全正确地检测出水印信息。

8.3.7 数字音频水印的评价标准

目前，对数字音频水印的评价尚无统一的标准。学术界和工业界提出了一些评估标准，其内容不尽相同。在研究过程中，一般选取水印系统最重要的三个指标来阐述水印系统的评估标准：不可听性(感知透明性)、鲁棒性和水印容量。这三者之间既相互依存又互为矛盾，一般来说，水印嵌入强度越大，则水印的鲁棒性越好，但同时水印的不可感知性就越差。如果要同时保持很强的鲁棒性和很好的不可感知性，就需要牺牲水印嵌入量。因此，实际应用中往往根据应用需求在三者之间找到一个适当的平衡。

1. IFPI 水印稳健性标准

国际留声机工业联盟(IFPI)在 1997 年对音频水印技术提出的稳健性要求可以看作是数字音频水印的最早标准。它要求水印标记应满足以下要求：

(1) 水印标记不能影响唱片的声音品质。

(2) 使用任何方法都不能删除或改变嵌入的信息，除非声音差到不能用的地步。

(3) 水印经过以下变换后必须能够恢复：各种滤波和信号处理操作(包括两个连续的 D/A 和 A/D 转换)；稳态压缩或 10%的时间扩张；压缩变换(如图像 MPEG 的数据压缩和多频带非线性振幅压缩)；添加加性或乘性噪声；使用同一系统加入另一个标记信号；使用低音和中音频段产生群时延失真或高达 15dB 的频率响应失真；群时延失真和陷波滤波等。

2. StirMark 标准

为了比较各种水印算法的优劣，应该有一个统一的测试标准。英国剑桥大学的 Fabien Petitcolas 等设计了一个通用的水印基准测试软件 StirMark，从 1997 年 11 月开始可免费下载并且公开了源代码。

StirMark 采用模块化设计以方便用户选择测试项目，测试的主要内容有感知性、算法容量、稳健性及速度等。

为了评价水印算法的健壮性，StirMark 根据攻击模式，提供了动态改变、滤波、回响、转换、有损压缩、添加噪声、调制、时域拉伸和基音改变及样点置乱等攻击方法。

3. 其他常见的评价方法

1）人耳的主观评价测试

向听音者提供三个信号：第一个是作为参照的原始信号，听音者知道它是原始信号；余下两个可能是原始信号，也可能是受攻击信号，对听音者是盲的。听音者对余下两个信号进行打分，分值为 1.0～5.0，分别代表从非常差到感知不到改变的音质，采用的评分标准是 ITV-R 制定的 5 分衰退等级。

由于该测试受人的主观因素影响较大，适合于作定性分析的场合。

2）信噪比

信噪比（Signal Noise Ration，SNR）可对水印算法本身引起的信号失真量进行定量评价。信噪比的定义如下：

$$\mathrm{SNR} = -10\lg\frac{\sum_i |A_i - A_i'|^2}{\sum_i |A_i|^2} \tag{8-1}$$

其中，A_i 为嵌入前的音频，A_i'为嵌入水印后的音频。

此外，还有水印算法采用峰值信噪比（Peak Signal-to-Noise Ratio，PSNR）对信号失真量进行定量评价。

3）比特错误率

比特错误率（Bit Error Rate，BER）在水印评价中也有应用，其定义如下：

$$\mathrm{BER} = \frac{\text{错误的比特数}}{\text{总比特数}} \times 100\% \tag{8-2}$$

4）归一化相关系数

在音频水印评价中，可以采用归一化相关系数（Normalized Cross-correlation，NC）定量地评价正在提取的水印与原始水印的相似性，归一化相关系数定义如下：

$$\mathrm{NC}(W,W') = \frac{\sum_{i=1}^{m1}\sum_{j=1}^{m2} w(i,j)w'(i,j)}{\sqrt{\sum_{i=1}^{m1}\sum_{j=1}^{m2} w(i,j)^2}\sqrt{\sum_{i=1}^{m1}\sum_{j=1}^{m2} w'(i,j)^2}} \tag{8-3}$$

其中，W、W'分别为原始水印序列、提取的水印序列。是否存在水印的判断标准是：

NC(W,W')$>T_c$,其中,T_c 为一阈值,其取值在0～1之间,T_c 的取值通常为0.5。

5) 水印容量

水印容量是指单位长度的音频中可以嵌入的水印信息量,通常用比特率 bps(bits per second)来表示。有的文献也以每千个采样样本中可嵌入的比特数来衡量。根据国际留声机联盟 IPPI 的要求,嵌入的水印信息量至少要到达 20bps 的带宽。

8.3.8 数字音频水印的发展趋势

数字音频水印的发展趋势主要体现在以下几方面。

1. 基于内容的水印技术研究

基于内容的水印技术强调将水印信息嵌入到音频信号的重要特征上。随着音频压缩标准的发展,数字音频水印方案应将水印与音频内容相结合,强调水印的同步性。基于音频内容或基于音频对象属性并与 HAS 相结合的水印方案将是数字音频水印的主要发展方向。

2. 结合各领域的先进思想进行水印技术研究

对现有数字音频水印算法的鲁棒性、数据率、感知等特性进行研究,结合数字信号处理技术,优化它们之间的关系,将各个领域的先进思想,如神经网络、模糊集、扩频、小波包和同步编码理论等融合进来,更好地发挥现有技术的优越性,创造更完美的水印技术。

3. 结合音频压缩标准与音频文件格式进行水印技术研究

现有的数字音频水印算法,对算法的研究很多,较少结合 WAV、MP3、MPEG、AC-3 等具体特性,同播放器和具体的网络协议相结合的研究也较少。结合具体音频压缩标准与文件格式,研究满足不同硬件和软件要求的水印算法,对数字音频水印技术的广泛应用具有重要的意义。

8.3.9 音频隐写术与数字水印的区别

数字音频隐写术和数字水印同属音频信息隐藏的范畴,它们都应用信号处理和编码等技术在数字音频信号中嵌入不可察觉的秘密信息,有许多共性和密切关系,但两者之间存在下列重要的差异。

1. 通信内容

数字水印的通信内容是音频信号本身,由数字水印提供对音频内容的版权保护;隐写术的通信内容是被隐藏的秘密信息,由音频信号提供对通信内容的安全掩护,音频信号的选取带有一定的任意性,只要不易引起人们的特别注意,并保证嵌入信息具有感官隐蔽性和统计隐蔽性。

2. 稳健性

数字水印必须高度保密，任何删除水印的操作都会损害数字音频产品的质量使之失去使用价值。隐写术则不一定要求这么强的稳健性，甚至很脆弱，修改或处理携密音频信号将破坏秘密信息使之不能被提取或者提取出错误的信息，从而阻止隐蔽通信。

3. 隐蔽性

数字水印的隐蔽性主要是指不影响音频信号的视听效果（商业价值），其存在性往往可以公布，公布申请音频产品受到数字水印的保护。而隐写术必须毫不引起局外人的注意，一旦秘密信息的存在性被察觉，即使内容未被破译，隐写亦告失败。对隐写术的分析常以揭示秘密信息的存在性为首要目标。

4. 嵌入数据量

数字水印通常只需要携带有关版权的少量信息，一些早期技术采用伪随机序列作为水印，用相关检测判断是否有某一特定的水印标记，实际上只有嵌入了 1 比特的信息，后来出现了大量多比特水印技术，但其数据量与大多数隐写应用相比仍然相差甚远。隐写术则不同，因为要实现隐蔽通信，往往要求携带足够数量的秘密数据。

8.4 数字音频隐写分析技术

目前对图像隐写分析的研究取得了不少进展。一般来说，对音频信号中的隐蔽信息的盲分析难度比图像中的隐写分析更高，目前在这方面的成果较少。基于音频的隐写分析方法的基本思想是，隐写术一定程度上无可避免地改变信号的统计特性，利用统计特性的差异设计通用分类器来区分载体信号与隐秘信号。

8.4.1 隐写分析原理

隐写分析是一件十分困难的工作，因为隐写分析者一般情况下只能获得隐写体，而对载体、嵌入算法、嵌入位置、嵌入密钥、加密密钥等信息一无所知。目前隐写分析的目标还只能是检测出媒体中是否含有秘密数据，如果能够估计出嵌入秘密数据的长度，就已经是比较高的水平。要想成功提取出秘密数据，除了已经判定所用的隐写工具，然后进行暴力破解，其他还未见先例。数字隐写的一般框架如图 8-4 所示。

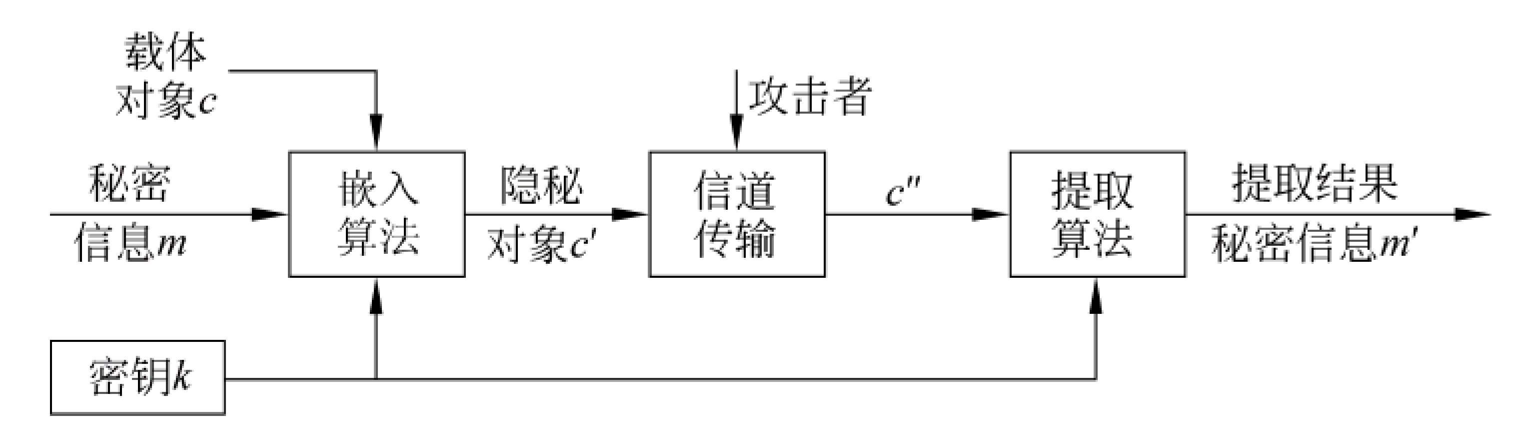


图 8-4 数字隐写的一般框架

由于隐写者必须通过修改原始数据才能实现秘密信息的嵌入，因而载体数据的统计特性不可避免地会发生一些变化。虽然分析者并不知道原始数据，但可以利用载体数据统计特性的异常来觉察到秘密信息的存在。从而可以给出如下定义，隐写分析是利用各种统计分析方法，揭示载体信号中隐蔽信息的存在性的技术。尽管并未破解秘密信息的具体内容，信道监控者还是可以阻断隐蔽通信并追查秘密信息的收发双方，导致隐写行为的失败。隐写分析是隐写术的主要威胁，因为一旦分析成功，隐写者不仅无法传送秘密信息，甚至有暴露身份的危险。

8.4.2 数字音频隐写分析分类

目前，数字音频隐写分析方法可分为基于感官检测的分析法、基于统计的分析法、基于特征的分析法和基于音频质量的分析法等。

1. 感官检测分析法

为了能够抵抗攻击，一般在载体比较敏感的区域隐藏信息，但同时也可能产生感官痕迹，从而暴露隐藏信息。感官检测利用人类感知和清晰分辨噪音的能力来对数字载体进行分析检测。在数字载体的失真和噪声中，人类可感知的失真或模式最易被检测到。辨别这种模式的一个方法是比较原始载体和隐密载体，注意可见的差异。如果没有原始载体，这种噪声就会作为载体的一个有机部分而不被注意。感官检测的思想是移去载体信息部分，这时人的感官就能区分剩余部分是否有潜在的信息或仍然是载体的内容。

当然，因为人的感知有一定的冗余度，且隐写算法的首要任务是不能超出人类视/听觉冗余度，人类感官系统不易察觉到隐蔽信息的存在，但这种变形和降质确实存在，可以配合对载体的处理，使得感官检测达到一定的功效。较为典型的对图像载体的处理手段是空间域图像位平面法，提取并显示图像的 LSB 平面，使得感官上更容易判断出差异模式，从而确定隐蔽信息的存在。感官检测不适合计算机的自动化分析检测，尤其是分析的媒体来源于网络，要求设计的分析算法必须满足实时性和低漏警率。

2. 统计检测分析法

这种分析方法是将原始载体的理论期望频率分布从可能是隐密的载体中检测的样本分布进行比较，从而找出差别的一种检测方法。信息隐藏只改变载体数据流的冗余部分，不改变感觉效果，但是经常会改变原始载体数据的统计性质。通过判定给定载体的统计性质是否属于正常情况，可以判断是否含有隐藏信息。

统计分析的关键是如何得到原始载体数据的期望频率分布，在大多数应用情况下，我们无法得到原始信号的频率分布，因为基于不同格式载体的信息隐藏方法多种多样，所以对它们进行统计攻击的具体方法也不尽相同。

3. 特征分析法

特征分析是以信息隐藏操作对载体造成的变化作为特征进行检测。这种特征可以是感官的、统计的或可以度量的。广义地来说，进行分析所依赖的就是特征，这种特征必

须根据具体的应用情况通过分析发现，进而利用这些特征进行分析。感官上的、格式上的特征一般来说较明显，也较容易分析，如基于文件格式中空余空间的信息隐藏分析，磁盘上未使用区域的信息隐藏分析等。

其他较复杂的隐藏特征则要根据隐藏算法进行数学推理分析，确定原始载体和隐密载体的度量特征差异，通过度量特征的差异分析信息隐藏。

4. 基于音频质量的分析方法

当秘密数据隐藏到音频文件中后，必然会引起音频质量的下降。HamzaOzer 等在 2003 年提出了一种基于音频质量和分类器的音频隐写分析方法。他们对表 8-1 中的 19 种有关音频质量的指标度量进行了调查：按照加性噪声模型，对待测音频进行小波去噪得到估计的原始音频，计算待测音频与小波去噪后的音频的各项指标的改变量，发现不含秘密数据的音频载体与包含秘密数据的音频隐写载体的各项指标的改变是不同的。

表 8-1 设计隐写分析器时测试的各项音频质量指标

感知域指标	非感知域指标	
	时 域 指 标	频 域 指 标
Bark Spectral Distortion(BSD)	Signal-to-noise ration(SNR)	Log-Likelihood ration(LLR)
Enhamed Modified Bark Spectral Distortion(EMBSD)	Czenakowski distance(CZD)	Itakura-Saito distance(ISD)
Perceptual Speech Quality Measure(PSQM)	—	COSH distance(COSH)
Perceptual Audio Quality Measure(PAQM)	—	Cepstral distance(CD)
Measuring Normalizing Block1(MNB1)	—	Short-Time Fourier-Radon Transform distance(STFRT)
Measuring Normalizing Block2(MNB2)	—	Spectral Phase Distortion(SP)
Weighted Slope Spectral distance(WSS)	—	Spectral Phase-Magnitude Distortion (SPM)
Modified Bark Spectral Distortion(MBSD)	Segmental Signal-to noise ration(SNRSog)	Log-Area ration(LAR)

对不同的隐写算法，会有不同的音频质量指标来准确、一致、单调地反映秘密数据的存在与数量。有学者采用方差分析(analysis of variance)和 SFS(Sequential Floating Search Method，顺序浮动搜索算法)两种方法分别对直接序列扩展频谱算法、跳频扩展频谱算法、回声算法、基于人的听觉系统的离散余弦算法、隐写软件 Stefanos 和 Stools 的鉴别音频质量指标进行了选择。然后又构造基于多元回归和支持向量机的两种分类器，作为通用的音频隐写分析器。隐写分析的基本流程如图 8-5 所示。

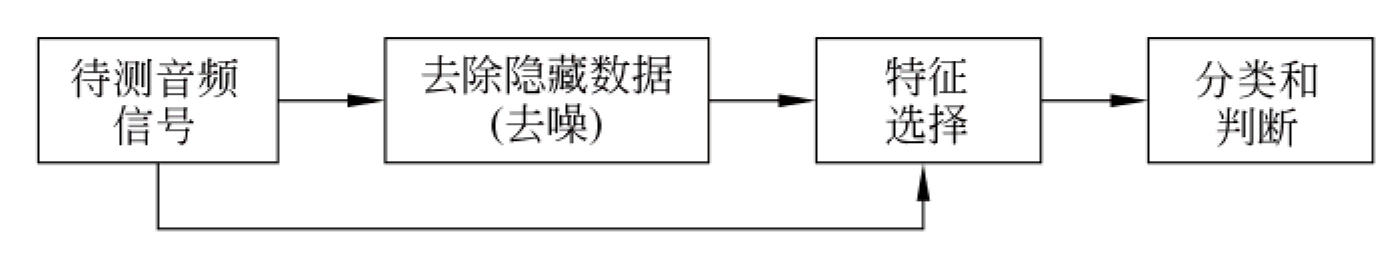


图 8-5 隐写分析流程图

8.4.3 隐写分析常用算法

隐写分析的研究途径主要有两类：一是针对某种具体的嵌入算法或软件进行研究，找出其固有的安全缺陷，从而实现对其可靠检测，称为专用隐写分析技术，如接下来将要介绍的 Chi-Square、RS 分析方法等；二是在分析隐写对载体所产生影响的基础上，找出对隐写敏感的低阶或高阶统计量，通过学习训练模型判断是否存在秘密信息，这种方法称为通用隐写分析技术。专用隐写分析技术可以准确检测采用某种嵌入方法的隐秘对象，准确性高但适用性低；通用隐写分析技术在整体上准确性也许不如专用隐写分析技术，但适用性高。

1. 专用隐写分析技术

专用隐写分析技术针对某种具体的嵌入算法或软件进行研究。针对数字音频隐写分析比较常见的专用隐写分析技术包括 Chi-Square 分析和 RS 分析。

1) Chi-Square 分析

卡方检验(Chi-Square test)是一种统计攻击的方法，该方法主要针对采用连续嵌入且嵌入信息服从均匀分布的 LSB 隐写方法。统计攻击的思想就是把隐秘对象的理论期望频率分布和从可能被修改的载体中观察到样本分布进行比较，从而找出差异，检测是否有信息嵌入。因为进行的是盲检测，没有原始载体作为比较，因此统计攻击的关键是如何得到理论频率分布。在隐写分析中，卡方检验统计测试的代表是 Povs(Pairs of Values)方法。

设一段音频样本值为 j 的出现频率为 n_j，其中，$j\in[0,255]$或 $j\in[0,65535]$。LSB 算法通常直接将音频样本值的最后一位用秘密信息取代，也就是说，如果秘密信息位与隐藏该位的样本值最后一位相同，则不改变原始载体；反之，则要改变样本值的最后一位，即将 $2i$ 改为 $2i+1$，或将 $2i+1$ 改为 $2i$，而不会将 $2i$ 改为 $2i-1$ 或将 $2i+1$ 改成 $2i+2$。LSB 隐写会改变样本值直方图，但因为样本值要么不变，要么在 n_{2i} 与 n_{2i+1} 之间互变，所以不会改变 $n_{2i}+n_{2i+1}$ 的值。秘密信息在嵌入之前往往经过加密，可以看作是 0、1 随机分布的比特流，而且值为 0 或 1 的可能性都是 1/2。如果秘密信息完全替代了载体音频的最低位，那么 n_{2i} 与 n_{2i+1} 的值会比较接近；而如果载体音频未经隐写，n_{2i}、n_{2i+1} 的值会相差较大。由上述可知，可以从随机样本中取得隐秘对象的理论期望频率分布。

令 $n_{2i}^*=\frac{n_{2i}+n_{2i+1}}{2}$，$q=\frac{n_{2i}-n_{2i+1}}{2}$，从理论上可以将 n_{2i}^* 作为期望频率分布。如果某个样本值为 $2i$，它对参数 q 的贡献为 1/2；如果值为 $2i+1$，对参数 q 的贡献为 $-1/2$。载体信号中共有 $2n_{2i}^*$ 个样本值为 $2i$ 或 $2i+1$，若所有样本都负载了 1 比特秘密信息，那么每个

样本值为 $2i$ 或 $2i+1$ 的概率为 0.5。当 $2n_{2i}^*$ 较大时，根据中心极限定理

$$\frac{n_{2i}-n_{2i+1}}{2\sqrt{n_{2i}^*}}=\frac{n_{2i}-n_{2i}^*}{\sqrt{n_{2i}^*}}\sim N(0,1) \tag{8-4}$$

其中$\sim N(0,1)$表示服从标准正态分布。因此构造出的统计函数为：

$$r=\sum_{i=1}^{k}\frac{(n_{2i}-n_{2i}^*)^2}{n_{2i}^*} \tag{8-5}$$

服从 χ^2 分布，自由度为 $k-1$。在式(8-5)中，k 等于 n_{2i}、n_{2i+1} 所组成数字对的数量，n_{2i}^* 小于 4 的情况不计在内，否则结果偏差会较大。r 越小表示载体含有秘密信息的可能性越大。结合 χ^2 分布的密度函数计算载体被隐写的可能性：

$$p=1-\frac{1}{2^{\frac{k-1}{2}}\Gamma\left(\frac{k-1}{2}\right)}\int_0^1\exp\left(-\frac{t}{2}\right)t^{\frac{k-1}{2}}\mathrm{d}t \tag{8-6}$$

如果 p 接近 1，则说明载体音频中含有秘密信息；如果 p 近似等于 0，则认为是原始载体。

2）Chi-Square 分析方法的扩展应用

LSB 算法不仅仅能在最低位平面嵌入信息，还可以通过改变最低几个位平面来达到秘密信息隐藏的目的。典型的信息隐藏软件 Hide4PGP 就可以在 16 位量化的声音中嵌入 4 比特的秘密信息。以最低位平面嵌入为例，利用 Chi-Square 分析方法揭示载体信号中隐蔽信息的存在性。Chi-Square 分析方法可以进一步推广到非最低位平面的 LSB 嵌入算法隐写分析中。

首先研究在最低 L 位平面上嵌入秘密信息的情况。即当量化精度为 N 的音频信号样本值 $X(i)$可以用 $X'(i)=[x_1(i),x_2(i),\cdots,x_L(i),\cdots,x_N(i)]$表示时，通过修改 $x_L(i)$来达到信息隐藏的目的。其中，$x_L(i)$为 $X(i)$第 L 位平面的值，$x_i(i)$为 $X(i)$最低位平面的值。

将每个样本值都除以 2^{L-1}，进行取整运算，即 $\hat{X}(i)=\lfloor X(i)/2^{L-1}\rfloor$，$\lfloor . \rfloor$为向下取整运算。然后利用 1)中的方法对 $\hat{X}(i)$进行分析，可以达到区分隐秘载体和原始载体的目的。在此，我们利用连续嵌入算法分别在语音中的第三位、第四位平面上嵌入随机的秘密信息。

3）RS 分析

RS 分析方法考虑的是图像各个位平面之间具有一定的非线性相关性，而当利用 LSB 隐藏信息后，这种相关性将被破坏。只要能找出衡量这一相关性的方法，并对隐藏信息前后的情况加以对比，就有可能设计出隐写分析方法。与图像类似，音频信号同样具有空间相关性。因此，可以基于 RS 研究音频载体的隐写分析方法。

首先将音频分割为相互独立的组，每个组 $G=(x_1,x_2,\cdots,x_n)$包含 n 个相邻的样本。例如，当选择 $n=4$，把 4 个连续的样本值作为一组，若音频信号采样位数为 8 位，则信号的样本值 $x_i\in P$，其中，$P=\{0,1,\cdots,255\}$。可以定义一个分辨函数，利用该函数描述每一组数据的随机程度。例如选择

$$f(x_1,x_2,\cdots,x_n)=\sum_{i=1}^{n-1}|x_i-x_{i+1}| \tag{8-7}$$

作为分辨函数。f 值越小，表明音频相邻值之间的起伏越小，而音频块的空间相关性越强。

定义二轮置换函数，即

$$F^2(x) = F(F(x)) = x, \quad x \in P \tag{8-8}$$

下面定义两种置换操作：F_1：$0\leftrightarrow1, 2\leftrightarrow3, \cdots, 254\leftrightarrow255$，即为 $2i$ 与 $2i+1$ 的相互关系；F_{-1}：$-1\leftrightarrow0, 1\leftrightarrow2, \cdots, 255\leftrightarrow256$，为 $2i-1$ 与 $2i$ 的相互关系，可表示为

$$F_{-1} = F_1(x+1) - 1 \tag{8-9}$$

类似地，定义 $F_0(x)$函数表示不变关系

$$F_0(x) = x \tag{8-10}$$

于是 LSB 隐写可表述如下：当待嵌入的秘密比特与样本值的 LSB 相同时，不作改动，即应用 F_0；不同时，应用 F_1 改变样本值。

利用函数 f 和置换 F 定义 R、S 和 U 三种类型的样本值组。

常规组：
$$G \in R \Leftrightarrow f(F(G)) > f(G) \tag{8-11}$$

异常组：
$$G \in S \Leftrightarrow f(F(G)) < f(G) \tag{8-12}$$

不变组：
$$G \in U \Leftrightarrow f(F(G)) = f(G) \tag{8-13}$$

对数据块的每个样本值应用翻转函数，记为

$$F(G) = (F_{M(1)}(x_1), F_{M(2)}(x_2), \cdots, F_{M(n)}(x_n)) \tag{8-14}$$

其中，$M(1), M(2), \cdots, M(n)$为 1，0 或 -1。置换操作 F 的功能是小范围置换样本值，从而模拟噪声的加入。对于原始语音，加入噪声会引起分辨函数 f 值的增加，从而使常规组的总数大于异常组的总数。

将待检信号分为很多大小相等的小块，对每个小块应用非负翻转，即 $M(1), M(2), \cdots, M(n)$为 1 或 0，常规组的个数记为 R_1，异常组的个数记为 S_1，同理，对于非正翻转，即 $M(1), M(2), \cdots, M(N)$为 -1 或 0，分别记为 R_2 和 S_2。

如果待检信号未经过 LSB 隐写处理，那么无论应用非负翻转还是非正翻转，从统计上来说，会同等程度地增加音频块的混乱度，也就是说，

$$\begin{cases} R_1 \approx R_2 \\ S_1 \approx S_2 \end{cases} \tag{8-15}$$

但是嵌入信息后，上式就不再成立。

定义隐写嵌入率为平均每个样本中含有秘密信息的比特数。比如，对于最大容量嵌入的最低位 LSB 算法来说，其隐写嵌入率为 1。随着嵌入率的增加，隐写行为使所有 LSB 逐渐随机化了，在此基础上再进行非负翻转并不能增加音频的空间相关性，即 R_1、S_1 的差距会随着嵌入率的上升而下降. 而对含密音频进行非正翻转，会有一些样本值经历两次翻转，但部分经历的是一次 F_1 翻转和一次 F_{-1} 翻转，与原始值之间就会偏离得很远。也就是说，两次翻转并不会互相抵消，所以 R_2、S_2 之间的距离不会随着嵌入率的上升而下降。

2. 通用盲隐写检测技术

通用检测算法主要是找出音频中对隐写敏感的低阶或高阶统计量，通过学习训练模型

判断是否存在秘密信息，由于通用检测方法较为复杂，下面简单介绍两种通用检测算法。

1）基于短时傅里叶变换和神经网络的隐写分析方案

在语音信号中，短时傅里叶变换（STFT）是目前最常用和最有效的时频信号分析处理法。可以利用 STFT 建立用于区分正常载体和隐秘载体的高阶统计特征模型。有学者提出了利用 STFT 和 PCA 提取音频信号的高阶统计特性和非支持向量机进行分类判决的方案，并给出了对 LSB 算法及 Hide4PGP 的隐写分析结果，如需了解具体算法可参考文献[55]。

2）高阶离散小波系数的隐写分析方法

利用基于高阶小波分解的统计建模可以检测到隐蔽信息的存在性。这个统计模型既包含基本的变换系数，又包含最优线性预测分析求出的变换系数的偏差统计。高阶统计反映出自然音频的某些内在特性。更重要的是，这些统计特性在信息嵌入后会发生显著的改变。该方法能进一步测试音乐信号中隐蔽信息的存在性。

8.4.4 隐写分析方法评价

对数字音频隐写分析方法的评价，可以采用如下 4 个指标：准确性、适用性、实用性和复杂性。

准确性指检测的准确程度，是评价被动隐写分析方法最重要的指标，可采用虚警率、漏报率和全局检测率表示。虚警率是把原始载体误判为隐写载体的概率；漏报率是把隐写载体误判为原始载体的概率。全局检测率是全面衡量准确性的指标。

适用性指检测算法对不同嵌入算法的有效性，可由检测算法能够有效检测多少种、多少类隐写术或嵌入算法来衡量。

实用性指检测算法可实际应用的程度，可由现实条件允许与否、检测结果稳定与否、自动化程度和实时性等来衡量。

复杂性是针对检测算法本身而言的，可由检测算法实现所需要的资源开销、软硬件条件等来衡量。到目前为止，还没有人给出适用性、实用性和复杂性的定量度量，只能通过比较不同检测算法之间的实现情况和检测效果得出一个相对的结论。

通用性强的隐写分析算法准确率不高，准确率高的隐写分析算法针对性太强。因此所有的隐写分析算法往往都有各自的优点、局限性和适用范围。要提高隐写分析算法的准确率，同时又具有较宽的适用范围是当今隐写分析的研究方向。一种思路是对待测音频应用各种音频隐写分析方法，融合各种检测的结果，应用人工智能判断秘密数据的存在与否；还有一种思路是建立一个分析系统，在系统中调度不同的算法检测不同统计特性的音频，达到算法与音频的最佳匹配适用性，从而提高系统的检测准确性。

8.5 数字音频取证技术

近几年来，随着数字录音设备的普及，数字录音大有取代以前模拟录音的趋势。各种音频处理算法和软件的广泛应用，使得一般的用户能轻易地对数字录音进行篡改而不

留下痕迹，因此从录音中听到的未必就是真实的。一段录音中可能有一些重要的字词被删除或者来源于其他录音的内容。如果虚假的录音被滥用，必将引起一系列的问题，如涉及到法律真实性、数字作品的版权、个人隐私的保护等。因为检测音频真实性和完整性有着十分重要的意义，针对音频篡改的取证技术也应运而生，并迅速成为信息安全的重要研究内容。本节通过数字音频取证步骤、数字音频取证技术分类以及一些常见的数字音频取证算法等来介绍音频数字取证技术。

8.5.1 数字音频取证技术步骤

音频取证是一门复杂的取证科学，需要对给定的录音进行反复的听力测试和使用仪器检验。进行音频取证需要确定该录音是不是原始的，并且解释其中出现异常的地方，例如录音信号出现不连续等。音频取证的最终目标就是尽可能地估计录音是否为在某一特定时间和地点所发生事件的真实记录。

最早开始正式接受录音作为证据的案件是1958年出现在美国的McKeever案件。当时美国法院同时提出了录音能够作为证据的七个条件，而这七个条件一直沿用至今：

(1) 录音设备必须能够录制对话并提供证据。

(2) 操作人员必须能够熟练操作设备。

(3) 录音必须可靠和正确。

(4) 录音不能修改、增添和裁剪。

(5) 录音必须进行保护直到法庭呈现。

(6) 必须识别出录音中的说话人。

(7) 对话必须是在自愿和诚信的情况下录制，说话人没有被诱导或强迫。

然而，使得音频取证真正引起公众和取证界关注的事件是发生在1974年美国历史上著名的“水门事件”。由六个顶尖科学家组成的调查组通过分析信号幅度、磁带回放的电声信号等特性，发现长达18.5min录音内容被人为抹除。“水门事件”对模拟音频取证影响深远，目前很多对模拟录音取证分析，都还是基于当年调查“水门事件”使用的技术。直到后来还出现了文献报道，提供了如何利用编辑技术令录音证据失效的详细过程。

另外，音频界的权威组织音频工程协会(Audio Engineering Society，AES)对“真实性的录音”也在AES27-1996标准中作了严格定义：“录音必须是所宣称的声音事件的同步记录，并且与录制宣称的录音方式完全一致；录音不能包含任何无法解释的人工痕迹、改变、增添、裁剪或者编辑。”

对于“录音的真实性认证”过程，音频工程协会在AES43-2000标准中也给出了严格要求：“取证分析人员必须检查给定的原始录音是否与给定的原始录音设备匹配。取证分析人员呈现的调查结果必须科学地表明给定的原始录音设备的确录制了给定的原始录音，并且在录制的内容中没有发现存在篡改、擅自编辑、故意删除内容或材料的痕迹。”

有关录音取证的问题可以分为两部分：

第一部分是评估所记录的声音事件与已知事件或证词声明是否一致，具体包括以下三方面：

(1) 评估录音中说话人与被指控人是否一致。

（2）评估录音的日期和时间与证词声明是否一致。

（3）评估录音地点与证词声明是否一致。

第二部分是评估录音是否具有原始性，即排除录音曾受到任何形式篡改的可能性。关于这方面的取证主要集中在以下四个方面：

（1）该录音是原始版本还是拷贝？

（2）从录音生成起，该录音是否曾被编辑或修改？

（3）该录音是否从宣称的录音设备获得？

（4）录音的内容是否与证词声明符合？

除了国际标准，我国在法律法规上对录音证据也有明确的规定。《民事诉讼法》第六十九条规定，人民法院对视听资料，应当辨别真伪，并结合本案的其他证据，审查确定能否作为认定事实的证据。从 2002 年 4 月 1 日施行的我国《最高人民法院关于民事诉讼证据的若干规定》对视听资料作为证据作了明确规定。具有证据效力的视听资料，必须同时符合以下三个条件：

（1）具备合法性。这要求音频证据的获得必须是通过合法手段，不能违反法律的禁止性规定。

（2）视听资料必须无疑点，即具备真实性。想把视听证据作为判案依据时，还要对视听证据是否有疑点进行审查。视听资料无疑点即具备真实性，就是当事人出示的视听证据未被裁接、剪辑或者伪造，前后连接紧密，内容未被篡改，具有客观真实性和连贯性。

（3）有其他证据佐证。在同时符合以上两个条件，且对方未提出反驳或反驳理由不成立时，法院应当确认视听证据的证明力。

8.5.2 数字音频取证的分类

音频取证技术主要分为三方面内容：说话人识别、语音增强和真实性认证。

1. 说话人识别

说话人识别或者嗓音比较从 20 世纪 60 年代开始应用，主要是为了解决司法系统上要求识别录音里面的说话人的问题。说话识别的过程分为两大步骤：提取表征说话人的特性参数与模型匹配。

提取说话人特征参数方面，常有的特征参数有短时傅里叶谱、基音频率、共振峰频率、线性预测系数、对数域比例等。随着使用单一特征参数在识别上准确率遇到的瓶颈，相关的研究方向开始转向使用混合特征参数。

在模型匹配方面，主要的方法有模板匹配、矢量量化、多层感知器、时延神经网络、混合高斯模型、最近相邻聚类和隐马尔科夫模型。

2. 语音增强

语音增强的任务是强化录音媒介上的衰弱信号，提高语音信号的清晰度。造成这种衰弱或者低音质的原因有很多，包括环境引起的卷积性衰弱、其他声音的掩蔽、电子噪声、低质量的录音系统等。常用的消除噪声技术分为时域处理方法和频域处理方法。时

域的经典算法是使用电平检测器，而频域最常用的技术就是频减技术。针对噪声的类型不同，研究人员分别提出了各种针对型的降噪方法。对于宽带噪声，最常用的方法是通过频减技术削弱；对于窄带噪声，可以利用自适应滤波器进行消减；脉冲噪声通常使用中值滤波的方法消除；卷积性噪声则利用倒谱均减技术削弱。

3. 真实性认证

从 20 世纪 70 年代至今，取证所用的语音信号通常来自于模拟磁带。对模拟磁带录音的认证包括以下几方面的工作：

(1) 物理检查：检查模拟磁带的长度、卷轴和外壳的状态，寻找外壳曾被打开或者磁带被拼接的痕迹。

(2) 监听异常：取证人员小心监听全部录音内容，标注所有明显的修改和异常处。记录下任何听得到的编辑和拼接痕迹、背景噪声或者磁带的不一致。

(3) 磁带特性：采用磁性处理技术，将该磁带和来自相同录音设备的参考信号进行比较，检验该磁带特性。

(4) 时域分析：观察录音信号的波形，其中录音起始和终止的过渡波形可以反映录音机的型号。

(5) 语谱图分析：利用语谱图分析器或者软件包检查语音信号不连续的痕迹，若该不连续是由非设备因素引起的，则判定该录音曾被修改。

8.5.3 数字音频取证常用算法

1. 基于电网频率的分析的算法

通常在录音时录音设备不仅捕获了语音信号，也记录了当时的电网频率(50 或 60Hz)。通过提取出的录音里面的电网频率，就能检测录音内容的完整性和认证录音的时间。

2. 基于检测录音设备环境的算法

早在 2005 年就有学者提出了一种利用音频分类和分析来确定说话环境真实性的算法框架，此方法可用于分类的四层次的音频特征(包括显著的句法特征、语义特征等)。2007 年有学者利用已知的音频隐写分析特征，提出一种机器学习的方法对录音地点和麦克风进行分类，实现根据录音判定录音的地点和麦克风。之后有学者利用录音设备的本地噪声鉴定音频文件真实性的方法。录音设备的本地噪声通过对由该设备产生的录音进行小波去噪提取。将待检测音频的本地噪声与录音设备的本地噪声比较，从而判断该音频文件是否包含由其他录音设备录制的视频。

3. 基于篡改痕迹的分析算法

由于自然语音信号在频域的高阶统计量具有弱相关性，而篡改的语音会大大增强这一相关性，有学者提出一种利用双边谱(bispectral)分析的方法检测语音的篡改。若篡改

后语音不作后期处理(如边界样本调整、压缩等),这种方法能准确地定位发生在WAV格式的语音信号的篡改。

之后有学者指出重采样的音频信号会引入周期性信息,并采用期望最大化对周期信息进行估计,实现检测音频是否重采样。

通过对插值音频信号频谱特性的分析,有学者发现插值后音频信号的各子带频谱的波动程度比正常音频的小。在此基础上,利用子带谱平滑度衡量音频信号在各子带频谱的波动程度,并借助子带谱平滑度检测音频信号的篡改。该方法可以检测出音频信号是否经过插值和拼接处理,同时还能估算出信号插值前的原始采样率。

实验证明,一段录音在生产过程中会产生失真和衰落。如房间的不同表面的多次反射,使得记录的声音在时域和频域都有拖尾效应。这些失真可以归结为音频回响时间。据此,有学者提出了一种对回响量进行建模和估计的方法,利用该回响量对录音进行取证。由于该研究还处在尝试阶段,实验结果估计的回响量与真实回响量还是有一定差异,但是足以检测出篡改的录音里面包含不同的回响量。

之后有学者提出了一种基于音频信号背景噪声的音频取证方法。在保证语音残余信号最小的情况下,提出了一种估计背景噪声的方法。在此基础上,利用基于相关度衡量检测语音信号的完整性。通过利用不同的语音信号在不同环境下进行录音实验,结果表明该方法比已知的语音算法具有更好的性能,同时获得更高的信噪比。

8.5.4 数字音频取证发展趋势

数字音频取证虽然取得了一些研究成果,但还是存在很多不足之处,尤其是以下几方面。

1. 篡改定位

目前音频取证需要专家反复辨听音频,确定可疑区域后再借助音频分析的方法定位篡改。当录音长度很长时,使用人工反复辨听是一项耗时耗力的工作,自动定位篡改可以大大加快取证进度。另外,定位篡改能够增强音频取证结果的说服力。

2. 针对压缩格式音频的分析

由于目前很多数字录音设备的录音文件格式均为MP3、WMA等压缩格式,而针对压缩格式音频的取证方法尚无报道。压缩格式音频的广泛性,使得对其篡改编辑也变得非常普遍。针对音频压缩过程中的痕迹,进行音频取证将发掘更多鉴别音频真实性的技术。

3. 鲁棒性

由于音频取证技术目前还处在研究阶段,有些算法还不能与实际的篡改检测情况相符。如在音频拼接检测中,目前的算法不允许拼接后进行处理操作。随着取证技术的发展,反取证技术也在不断进步,各种篡改后的后期处理技术不断地向现有的取证技术发出挑战。因此,研究数字音频取证技术的时候,必须将其对抗常有的后期处理技术的鲁棒性考虑在内。

尽管数字音频取证技术还处在起步阶段，但数字图像取证已经产生了一系列数字图像取证的方法，例如复制粘贴检测、JPEG重压缩检测、重采样检测、模式噪声检测、CFA插值检测、自然图像模型等。虽然音频和图像两种截然不同的媒介，导致很多合适图像取证的方法无法直接应用在音频上，但是图像取证的很多指导思想是值得音频取证借鉴的。例如利用压缩遗留的痕迹、检测背景噪声的一致性等。

MP3是当前最常见、使用最广泛的数字音频格式，无论网上的音乐格式还是数字录音笔保存文件的格式大多都为MP3格式，但针对这一格式的音频取证技术的报道还很少，因此很有必要开发与MP3相关的取证技术，包括鉴别假音质的音乐、鉴别音频的原始性、定位音频的篡改、录音与录音笔的匹配等。这些将可能成为数字音频取证的重要内容。

思　考　题

8.1　简述音频文件的常见存储格式及其各自特点。
8.2　常见的音频素材获取途径有哪几种？
8.3　动手制作一个数字音频作品，要求：
　　题材：翻唱歌曲，诗配乐，朗读，演讲，彩铃……
　　格式：MP3，位速320kbps
8.4　什么是采样？什么是量化？什么是编码？
8.5　论述回声的相关原理。
8.6　假设目前正在使用麦克风进行录音，采样频率设为22kHz，量化选为16位，在不采用压缩技术的情况下，计算录制57秒的立体声文件大约需要多少空间？
8.7　简述数字音频信号的一般处理过程。
8.8　分析数字音频完全加密和选择性加密的优缺点。
8.9　列举数字音频水印技术的用途(至少5例)。
8.10　数字音频水印与一般图像水印有何异同？
8.11　数字音频的频域水印算法有什么优点？
8.12　举例说明常见数字音频水印算法并分析其各自特点。
8.13　列出常见隐写方法并分析其特点。
8.14　查找一个数字音频隐写工具，做简单的隐写分析实验。

参 考 文 献

[1] Kuhn M G. Analysis of the aggravation video scrambling method. From http://www.el.cam.ac.uk//-mgk25.

[2] Tang T. Methods for encrypting and decrypting MPEG video data efficiently. In Proceedings of the 4th ACM International Multimedia Conference, Boston, MA, 1996, 219-230.

[3] Maples T, Spans G. Performance study of a selection encryption scheme for the Security of networked real-time video. In: proceedings of International conference on computer

Communications and Networks, Las Vegas, NV, September 20-23, 1995.
[4] 王卓,赵千川.基于能量量化的音频水印算法.计算机工程与应用,2004,(24):48-51.
[5] 李跃强,孙星明.一种健壮的数字音频水印时域算法.计算机工程与应用,2005,41(8):89-91.
[6] 阚言亮.基于傅里叶变换的数字音频水印技术.北京广播学院学报(自然科学版).2005 12(1):62-67.
[7] 王向阳,杨红颖.基于离散余弦变换的自适应数字音频水印技术研究.小型微型计算机系统,2004,25(10):1825-1827.
[8] 陈琦,张连海.一种基于特征点检测的音频数字水印算法.计算机应用,2004,24(6):198-200.
[9] 赵静,周明全.基于 DCT 变换及 SVD 处理的音频数字水印算法.微机发展,2005,15(2):50-52.
[10] 陈荔聪,姚志强.基于局部极值点的音频盲水印算法.计算机工程与应用,2005,41(19):41-43.
[11] 王向阳,杨红颖.一种可抵抗 Mp3 压缩的音频水印算法.自动化学报,2007,33(3):248-252.
[12] Xu C S, Zhu Y W, Feng D D. A Robust and Fast Watermarking Scheme for Compressed Audio. In: Proceedings of IEEE International Conference on Multimedia and Expo, 2001: 253-256.
[13] Qiao L. Narrated K. Non-Invertible Watermarking Methods for MPEG Encoded Audio. SPIE Proceedings on Security and Watermarking of Multimedia Contents, 1999, 194-202.
[14] Koukopoulos D, Stamatiou Y. An Efficient Watermarking Method for MP3 Audio Files. In Proceedings of IEC (Prague), 2005, 154-159.
[15] Takagi K, S. Sakazawa, Y. Takishima. Light Weight MP3 Watermarking Method for Mobile Terminals. In: Proceedings of the 13m Annual ACM International Conference on Multimedia, 2005, 443-446.
[16] Bender W, et a1. Techniques for data hiding. IBM System Journal, 1996, 35(3&4): 313-336.
[17] Gruhl D, Lu A, W. Bender. Echo hiding Information hiding. In Proceedings of the first International Workshop on Information Hiding, Cambridge, UK, 1996, 114: 295-315
[18] 赵朝阳,刘振华,王挺.数字音频信号的回声数据隐藏技术.计算机应用研究,2000,17(7):42-44
[19] Hyen O Oh, et a1. New echo embedding technique for robust and imperceptible audio watermarking. In: Proceedings of the Acoustics, Speech, and Signal Processing, 2001, 3: 1341-1344.
[20] Wang Y. A new watermarking method of digital audio content for copyright Protection. In Proceedings of ICSP 98, 1998, 1: 1420-1423.
[21] 钮心忻,杨义先.基于小波变换的数字水印隐藏与检测算法.计算机学报,2000,23(1):21-27.
[22] 陈琦,王炳锡.一种基于 DCT 变换的语音数字水印算法研究.信号处理,2001,17(3):238-241.
[23] Neubauer C, Here J. Audio watermarking of MPEG-2 AAC bit streams. In: Proceeding s of the 10th Audio Engineering Society Convention, 2000, Paris France.
[24] Siebenhaar F, et al. New results on combined audio compression/Watermarking. In: Proceedings of the 11th Audio Engineering Society Convention, 2001, New York USA.
[25] 马田,张新鹏,王朔中.数字音频信号中的频域扰动调制水印嵌入.信号处理,2002,18(3):202-207.
[26] 王泳,黄继武,Shi Y Q.快速重同步的有意义音频水印盲检测算法.计算机研究与发展。2003,40(2):215-230.
[27] 张开文,张新鹏,王朔中.图像及音频信号中隐蔽嵌入信息存在性的统计检验.电子与信息学报,2003,25(7):872-877
[28] Bender W, Grubl D, Morimoto N. Techniques for Data Hiding. IBM Systems Journal, 1996, 35:

313-336.

[29] Bolt R H, Cooper F S, Flanagan J L, McKnight J G, et al. Report on a technical investigation conducted for the U. S District Court for the District of Columbia by the U. S Advisory Panel on White House Tapes. U. S Government Printing Office, Washington, D. C. 1974.

[30] Owen T. Forensics Audio and Video-Theory and Applications. Journal of the Audio Engineering Sociey, 1988, 36: 36-41.

[31] Warriner W. A Guide To Tape Splicing: How to Falsify Evidence and Other Diversions. High Fidelity Magazine, 1975: 48-53.

[32] AES43: 2000. Standard for Forensic Purposes-Criteria for Authentication of Analog Audio tape recordings. Audio Engineering Society, 2000.

[33] Fu D D, Shi Y Q, Su W. A generalized Bedford's law for JPEG coefficients and its application in image forensics. In Proceedings of SPIE on Security, Steganagraphy, and Watermarking of Multimedia Contents, San Jose, 2007, 6505: 47-58.

[34] Broder A P A. Forensic Speech and Audio Analysis Linguistics: 1998 to 2001 A Review. In: Proceedings of 13th Interpol Forensic Science Symposium, Lyon, France, 2001: 54-84.

[35] Brixen E B. Techniques for the authentication of digital audio recording. In: Proceedings of 12th Conference on Audio Engineering Society, Vienne, Austria, 2007, Paper 7014.

[36] French J P. Development in Forensic speaker Identification. Institute of Acoustics, Acoustic cs Bulletin, 1993, 18(5): 13-16.

[37] Rabiner L R, Juang B. Fundamental of Speech Recognition Prentice Hall, Englewood Cliffs, NJ, 1993.

[38] Campbell J P. Speaker recognition a tutorial. Proceedings of the IEEE, 1997. 85(9): 1437-1462.

[39] Doddington G. A method of speaker verification. Journal of Acoustic Society American, 1971, 49(A): 139-143.

[40] Sambur M R. Speaker recognition and verification using linear prediction analysis. Phd's Thesis in MIT, 1972.

[41] Grigoras C. Digital Audio Recording Analysis: The Electric Network Frequency (ENF) Criterion. The International Journal of Speech Language and the law, 2005, 12(1): 63-76.

[42] Grigoras C. Applications of ENF criterion in forensic audio, video, computer and telecommunication analysis. Forensic Science International, 2007, 16(2): 136-145.

[43] Grigoras C. Application ENF Analysis in Forensic Authentication of Digital Audio and Video Recording. Journal of AES, 2009, 57(9): 643-661.

[44] W. Sanders R. Digital Authenticity Using the Electric Network Frequency. In: Proceedings of 33rd International Conference: Audio Forensics-Theory and Practice, Denver, 2008: 24-30.

[45] Nicolalde D P, Apolinario J A. Audio authenticity: Detecting ENF discontinuity with high precision phase analysis. IEEE Transactions on Information Forensics and Security, 2010, 53(3): 534-543.

[46] Buchholz R, Kraetzer C, Dittmamn J. Microphone Classification Using Fourier Coefficients. Information Hiding, Lecture Notes in Computer Science, 2009, 5806: 235-246.

[47] 邵松华，黄征，徐彻，等. 数字音频与录制设备的相关性研究. 计算机工程，2009，35(2): 224-226.

[48] Farid H. Detecting Digital Forgeries Using Bispectral Analysis. MIT AI Memo AIM-1657, MIT,

1999.
[49] 姚秋明,柴佩琪,宣国荣,等. 基于期望最大化算法的音频取证中的篡改检测. 计算机应用,2006,26(11):2598-2601.
[50] 丁琦,平西建. 基于子带谱平滑度的音频篡改检测. 应用科学学报,2010,28(2): 142-146.
[51] Malik H,Farid H. Audio forensics from acoustic reverberation. In Proceedings of ICASSP 2010, Dallas,2010:1710-1713.
[52] Bohme R, Westfeld A. Statistical Characterizations of MP3 Encodes for Steganalysis. In: Proceedings of 6th ACM Multimedia and Security Workshop,Magdeburg,Germany,2004:25-34.
[53] Lukas J,Fridrich J. Estimation of primary quantization matrix in double compressed jpeg images. In: Proceedings of Digital Forensic Research Workshop,Cleveland,2003:35-46.
[54] Johnson M K, Lyu S, Farid H. Steganalysis of Recorded Speech. In: Proceedis of SPIE, 2005, 5681:664-672.
[55] 杨正琴.音频隐写分析技术的研究.南京理工大学硕士学位论文,2006.
[56] 杨锐.数字音频取证技术研究.中山大学博士学位论文,2010.
[57] 王燕.语音隐写分析技术研究.华北电力大学硕士学位论文,2008.
[58] 张慧华.基于超混沌序列加密的数字音频水印算法.广东工业大学博士学位论文. 2009
[59] 杨帆. MIDI 音频隐写分析研究. 中国科技大学硕士学位论文. 2009.
[60] 要强.基于心理学模型的 AVS 音频水印算法研究.天津大学硕士学位论文. 2009.
[61] 何松.利用 HAS 掩蔽效应的变换域语音隐写算法研究与分析.苏州大学硕士学位论文,2006.
[62] http://baike. baidu. com/view/66015. htm.
[63] http://baike. baidu. com/view/8136. htm.
[64] 吴建军,杨格兰.信息隐藏与数字水印的研究与发展.山西科技,2006,1: 16-17.

第 9 章 chapter 9

数字视频内容安全

本章学习目标

随着数字视频产品的日益增多,数字视频的安全保护开始引起了人们的关注。本章将对数字视频内容安全的有关概念和方法进行介绍,主要包括数字视频内容加密、数字视频隐写与水印、数字视频隐写分析技术与数字视频取证等方面的知识。

通过本章的学习,应掌握以下内容:

(1) 数字视频压缩编码技术。

(2) 数字视频内容加密:完全加密、选择加密、混沌加密。

(3) 数字视频隐写与水印技术。

(4) 数字视频隐写分析技术。

(5) 数字视频取证技术。

9.1 数字视频内容安全基本概念

随着大量消费类数字视频产品在市场上的推出,如 VCD、DVD,以及网络视频分享网站的火爆,如 Youtube、Youku 等,使得视频作品能够很容易获得,制作其完美拷贝也变得非常容易。这给人们的工作、学习和生活带来了极大的便利,但同时也可能会导致大规模的非授权拷贝、秘密信息的泄露等一系列的安全问题,为了解决这些问题,本章将针对数字视频内容安全技术进行介绍。

9.1.1 数字视频概述

1. 数字视频的特点

数字视频就是以数字形式记录的视频,和模拟视频是相对的。数字视频就是通过数字摄像机等视频捕捉设备,将外界影像的亮度和颜色等信息转化为数字视频信号,然后存储在存储介质上(如光盘、磁盘等),得到不同格式的数字视频。播放时,视频信号通过视频播放器被转变为帧信息,并以每秒约 30 帧的速度投影到显示器上,让人的眼睛感觉它是连续不间断地运动着的。

和模拟视频信号相比,数字视频具有许多突出优点。

1）失真小、噪声低、视频质量高

模拟电视信号在放大、处理、传输、存储过程中，难免会引入失真和噪声，多种噪声与失真叠加到源信号上，不易去除，而且会随着处理次数和传输距离的增加而累积，导致图像质量及信噪比下降。相反，数字视频设有上述的噪声累积效应，只要噪声电平不超过信号脉冲幅度一半，就可对其整型，并恢复成 0 和 1 两种电平，不会引入噪声。

2）易处理、易校正

数字视频信号利用 VLSI 芯片进行压缩编码处理、彩色校正等处理相对来说容易得多。随着专用芯片和通用 DSP 的发展，视频数字压缩编码取得更大发展。

3）容量大、节目多

同样带宽容纳的数字电视节目比模拟的多得多。例如，CATV 频道中，550～750MHz 的 200MHz 带宽中，如果传送模拟电视，每个节目需 8MHz 带宽，最多传送 25 套节目。如果换成数字节目，采用 64QAM 调制，频谱利用率为 5.3，如果每路节目用 MPEG-2 压缩为 2Mb/s，实际只需 4/5.3≈0.75(MHz)带宽，于是在同样的 200MHz 带宽中可传送 200/0.75≈260 套节目，约为模拟电视的 11 倍。

2. 采样与量化

为了存储数字视频信号，模拟视频信号必须通过模拟/数字(A/D)转换器来转变为数字的 0 或 1。这个转变过程就是视频捕捉(或采集过程)。如果要在电视机上观看数字视频，则需要一个从数字到模拟的转换器将二进制信息解码成模拟信号，才能进行播放。

电视信号有两种采样：时间取样和空间采样。

1）时间取样

运动图像可由每秒若干帧的静止图像构成，我国采用 PAL 制规定彩色电视 25 帧/s，美国、日本等采用 NTSC 制规定 30 帧/s。

2）空间取样

在同一电视信号帧中，同一行由若干取样点构成，即像素，这种取样点就属于空间取样。例如，国际上标准电视格式为 720×576 像素，即每帧 576 行，每行 720 个像素。

不同的国家采用不同制式，为实现国际不同彩色电视制式国家间通信，通常采用一种公共格式(CIF)，如表 9-1 所示。

表 9-1　视频帧格式

格　式	亮度清晰度	格　式	亮度清晰度
亚 QCIF	96×128	CIF	288×352
QCIF	144×176	4CIF	576×720

3. 彩色空间

黑白图像的每个像素中只需一个幅值表示其亮度即可，而彩色图像的每个像素至少需要三个值表示其亮度和色度。

1）RGB

任何彩色图像可由不同比例的红、绿和蓝色组合而成，即三基色原理。

2）YCbCr(YUV)

人类视频系统(HVS)对亮度比彩色更敏感，因此可把亮度从彩色信息中分离出来，并使之具有更高的清晰度。

如果亮度分量用 Y 表示，色度用 Cb、Cr 表示，则由大量实验得出：

$$\begin{cases} Y = 0.299R + 0.857G + 0.114B \\ \mathrm{Cb} = 0.564(B - Y) \\ \mathrm{Cr} = 0.713(R - Y) \end{cases} \tag{9-1}$$

反之，可由下式得到相应的 R、G、B：

$$\begin{cases} R = Y + 1.402\mathrm{Cr} \\ G = Y - 0.344\mathrm{Cb} - 0.714\mathrm{Cr} \\ B = Y + 1.772\mathrm{Cb} \end{cases} \tag{9-2}$$

4. 连续视频取样格式

有三种不同的彩色视频取样格式，如图 9-1 所示。

(1) 4∶4∶4——Y、Cb 和 Cr 具有同样水平和垂直清晰度，每个像素位置都有 Y、Cb 和 Cr 分量。

(2) 4∶2∶2——彩色分量和亮度分量具有同样的垂直清晰度，水平方向上，每四个亮度像素有两个 Cb 和两个 Cr。

(3) 4∶2∶0——水平和垂直方向上，Cb 和 Cr 都是 Y 的一半。

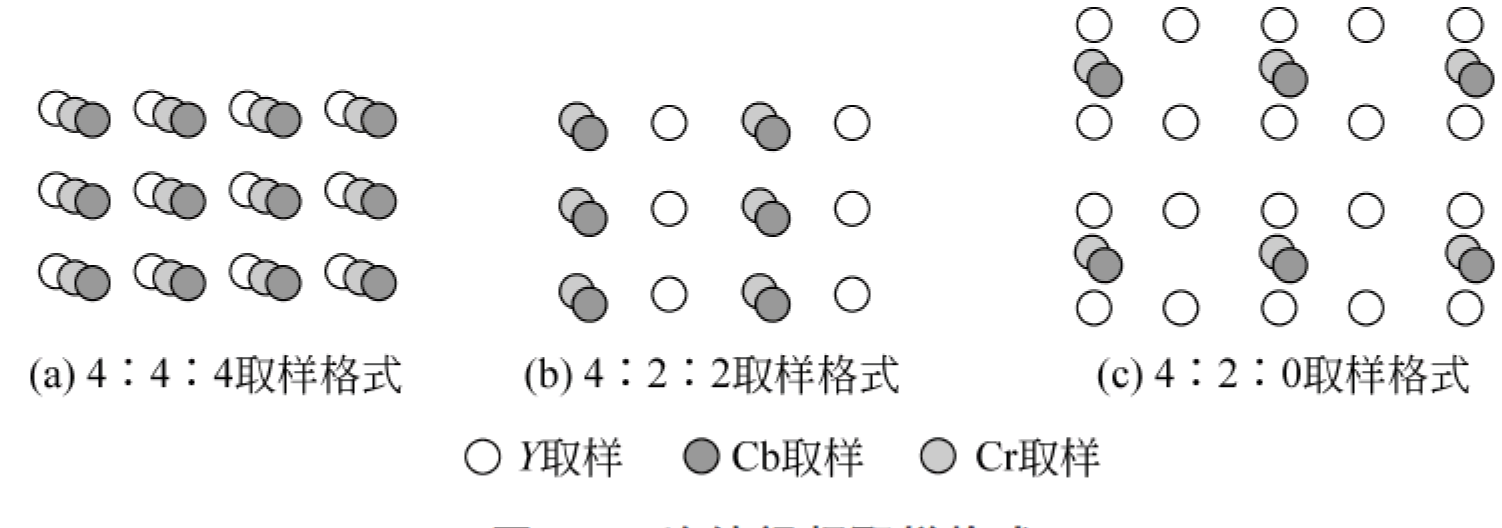


图 9-1 连续视频取样格式

9.1.2 数字视频压缩编码基础

1. 压缩的必要及可行性

视频文件的数据量十分庞大，给存储和传输带来很大不便。据计算，数字电视如果播放 ITU. R601 标准的未经压缩的视频，需要 216Mb/s 的传输带宽。如果按照这种码率，一张 4.7GB 的 DVD 光盘仅能存放 87s 的视频。可见当前的存储容量和网络带宽远不能满足视频数据高码率的需求，所以视频在存储或传输之前通常需要压缩。

实际上视频数据中存在大量的冗余信息：

1）空间冗余

空间冗余是指画面中相邻像素间或数个相邻像素块间存在高度的空间相关性。例

如在一幅蓝天白云图中，画面中绝大部分表示天空背景的像素都是蓝色的，白云的颜色也是相近的，在存储图像时，就不必存储每一点的数据，可以记录下图像的特征，重现画面。这种代码的数据量是很小的。因此说图像中相邻像素间或数个相邻像素块间存在高度相关性是空间冗余编码的依据。

2）时间冗余

数字视频的相邻帧往往包含相同的背景和移动物体，只不过移动物体所在空间位置会有微小的变动，这就产生了大量的数据冗余，称为时间冗余。这样就可以通过帧间预测、运动补偿以及运动估计等方法，获得运动矢量等极少量信息来表示下一帧图像，从而减少帧序列冗余信息。

3）心理视觉冗余

主要利用人类视觉系统对视觉信息的不同敏感度，把那些不十分重要的信息称作视觉冗余，如人眼对亮度信息比彩色更敏感，保证亮度信息时，可以把彩色信息清晰度降低，就可以显著压缩带宽，实现视频压缩的目的。

4）编码冗余

如果表示视频信息内容使用的平均比特数大于该消息的信息熵，则信源中存在冗余，即信息熵冗余。

2. 常用的压缩策略

为对视频信息进行压缩，可以用多种不同的方法和策略，下面是几种常用策略：

1）有损压缩

有损压缩会丢弃一些数据，以便获得较低位速。压缩的过程中要丢失一些人眼和人耳所不敏感的图像信息，而且丢失的信息不可恢复。几乎所有高压缩的算法都采用有损压缩，这样才能达到低数据率的目标。

2）无损压缩

无损压缩即压缩前和解压缩后的数据完全一致。因为在不丢失信息的前提下，无损压缩节省的空间较少。

3）心理视频压缩

心理视觉模型去掉的是眼睛不需要的视频数据。假设有一个在60秒未经压缩的视频片段，视频始终显示位于同一位置的物体，即每帧图像中都会重复这个物体的数据。如果使用了心理视觉压缩，就会把该物体在一帧图像中的数据存储下来，以在接下来的帧中使用，从而节省大量数据。

9.1.3 数字视频常见格式

数字视频通常包括运动的图像、声音背景音乐和音效，具有数据量大和实时性强等特点。针对不同的应用要求，有多种文件格式：

（1）RM格式：Networks公司所制定的音频视频压缩规范称为Real Media，它主要包含Real Audio、Real Video和Real Flash三部分，Real Media可以根据不同的网络传输速率制定出不同的压缩比率，从而实现在低速率的网络上进行影像数据实时传送和

播放。

(2) RMVB格式：这是一种由RM视频格式升级延伸出的新视频格式，它的先进之处在于RMVB视频格式打破了原先RM格式那种平均压缩采样的方式，在保证平均压缩比的基础上合理利用比特率资源，就是说静止和动作场面少的画面场景采用较低的编码速率，这样可以留出更多的带宽空间，而这些带宽会在出现快速运动的画面场景时被利用。它能在保证静止画面质量的前提下，大幅地提高运动图像的画面质量。

(3) ASF格式：英文全称为Advanced Streaming Format，它是微软公司为了和现在的Real Player竞争而推出的一种视频格式，用户可以直接使用Windows自带的Windows Media Player对其进行播放。使用了MPEG-4的压缩算法，所以压缩率和图像的质量都很不错。

(4) AVI格式：它的英文全称为Audio Video Interleaved，即音频视频交错格式，可以将视频和音频交织在一起进行同步播放。这种视频格式的优点是图像质量好，可以跨多个平台使用，但是其缺点是体积过于庞大，而且更加糟糕的是压缩标准不统一，因此经常会遇到高版本Windows媒体播放器播放不了采用早期编码编辑的AVI格式视频，而低版本Windows媒体播放器又播放不了采用最新编码编辑的AVI格式视频的情况。

(5) AMV格式：相对于MTV格式来说，AMV视频格式比MTV视频格式有着更好的压缩比例以及画面质量。

(6) SWF格式：利用Flash可以制作出一种后缀名为SWF(Shockwave Format)的动画，这种格式的动画图像能够用比较小的体积来表现丰富的多媒体形式。在图像的传输方面，不必等到文件全部下载才能观看，而是可以边下载边看。此外，SWF动画是基于矢量技术制作的，因此不管将画面放大多少倍，画面都不会因此而有任何损害。

(7) MPEG格式：MPEG是Motion Picture Experts Group的缩写，它包括MPEG-1、MPEG-2和MPEG-4(注意，没有MPEG-3，MP3只是MPEG Layer 3)。MPEG-1广泛应用在VCD的制作中，可以说99%的VCD都是用MPEG-1格式压缩的。MPEG-2则是应用在DVD的制作(压缩)方面，同时在一些HDTV(高清晰电视广播)和一些高要求视频编辑、处理上也有相当的应用面。MPEG-4是一种新的压缩算法，使用这种算法的ASF格式可以把一部120min长的电影(未视频文件)压缩到300MB左右的视频流，可供在网上观看。

(8) DivX格式：DivX由Microsoft mpeg4v3修改而来，使用MPEG-4压缩算法。MPEG4压缩一部DVD只需要2张CDROM。这样就意味着，用户不需要买DVD ROM也可以得到和它差不多的视频质量了。

(9) WMV格式：WMV(Windows Media Video)也是微软公司推出的一种采用独立编码方式并且可以直接在网上实时观看视频节目的文件压缩格式。WMV格式的主要优点包括本地或网络回放、可扩充的媒体类型、部件下载、可伸缩的媒体类型、流的优先级化、多语言支持、环境独立性、丰富的流间关系以及扩展性等。

(10) QuickTime：QuickTime(MOV)是Apple(苹果)公司创立的一种视频格式，支持MAC机和Windows平台，是一种优良的视频编码格式。

(11) RGB：对一种颜色进行编码的方法统称为“颜色空间”或“色域”。RGB(红、绿、

蓝)只是众多颜色空间的一种。采用这种编码方法,每种颜色都可用三个变量来表示:红色、绿色以及蓝色的强度。记录及显示彩色图像时,RGB 是最常见的一种方案。但是,它缺乏与早期黑白显示系统的良好兼容性。因此,电子电器厂商普遍采用的做法是,将 RGB 转换成 YUV 颜色空间,以维持兼容,再根据需要换回 RGB 格式,以便在计算机显示器上显示彩色图形。

(12) YUV:YUV(亦称 YCrCb)主要用于优化彩色视频信号的传输,使其向后兼容老式黑白电视。与 RGB 视频信号相比,它最大的优点在于只需占用极少的带宽(RGB 要求三个独立的视频信号同时传输)。

9.1.4 数字视频编码技术

目前数字视频压缩编码技术主要采用 MPEG-X、H.264/AVC、Real Video 等几种编码技术。对于用户而言,最关心的主要有清晰度、存储量、价格等,采用不同的压缩技术,将很大程度地影响以上因素。

1. MJPEG

MJPEG 是指 Motion JPEG(Motion Joint Photographic Experts Group),即动态 JPEG。它是由 JPEG 专家组制定的,它把视频序列看做连续的静止图像,不考虑视频流中不同帧之间的变化,只单独对某一帧进行压缩,通常可达到 6∶1 的压缩率。但由于没有考虑帧间变化,造成大量冗余信息被重复存储。因为每帧都可任意存取,所以 MJPEG 常用于视频编辑系统。动态 JPEG 能产生高质量、全屏、全运动的视频,但是,它需要依赖附加的硬件。而且,由于 MJPEG 不是一个标准化的格式,各厂家都有自己版本的 MJPEG,双方的文件无法互相识别。

MJPEG 的优点是画质比较清晰,缺点是压缩率低,占用带宽很大。一般单路占用带宽 2MB 左右。

2. MPEG-X

(1) MPEG-1,即 VCD 标准,制定于 1992 年,它用于传输 1.5Mb/s 数据传输率的数字存储媒体运动图像及其伴音的编码,经过 MPEG-1 标准压缩后,视频数据压缩率为 1/100~1/200,影视图像的分辨率为 360×240×30(NTSC 制)或 360×288×25(PAL 制),它的质量要比家用录像系统(Video Home System,VHS)的质量略高。音频压缩率为 1/6.5,声音接近于 CD-DA 的质量。MPEG-1 的编码速率最高可达 4-5Mb/s,但随着速率的提高,其解码后的图像质量有所降低。

(2) MPEG-2,即 DVD 标准,制定于 1994 年,传输速率在 3~10Mb/s,与 MPEG-1 兼容,适用于 1.5~60Mb/s 甚至更高的编码范围。分辨率为 720×480×30(NTSC 制)或 720×576×25(PAL 制)。MPEG-2 的音频编码可提供左、右、中和两个环绕声道,以及一个加重低音声道,和多达七个伴音声道(DVD 可有八种语言配音的原因)。MPEG-2 可提供一个较广的范围改变压缩比,以适应不同画面质量、存储容量以及带宽的要求。采用 MPEG-2 传输的视频的画质质量最好,但同时也需要非常大的带宽,通常在 4~

15MB。MPEG-2 不太适合远程传输。

(3) MPEG-4。如果说 MPEG-1"文件小,但质量差",而 MPEG-2 则"质量好,但更占空间"的话,那么 MPEG-4 则很好地结合了两者的优点。它于 1999 年 1 月成为一个国际性标准,它是超低码率运动图像和语言的压缩标准。MPEG-4 标准主要应用于视像电话(video phone),视像电子邮件(video E-mail)等,其传输速率要求较低,在 4800~64kb/s,分辨率为 176×144。MPEG-4 利用很窄的带宽,通过帧重建技术,压缩和传输数据,以求以最少的数据获得最佳的图像质量。MPEG-4 为多媒体数据压缩提供了一个更为广阔的平台。它主要提出格式、架构的定义,而不是具体的算法。它可以将各种各样的多媒体技术充分利用,包括压缩本身的一些工具、算法,也包括图像合成、语音合成等技术。

MPEG-4 的特点是其更适于交互式 AV 服务以及远程监控。MPEG-4 是第一个使用户由被动变为主动(不再只是观看,允许用户加入其中,即有交互性)的动态图像标准;它的另一个特点是综合性,从根源上说,MPEG-4 试图将自然物体与人造物体相融合(视觉效果意义上的)。MPEG-4 的设计目标还有更广的适应性和可扩展性。MPEG-4 标准的占用带宽可调,占用带宽与图像的清晰度成正比。

3. H.264/AVC

H.264 是 MPEG-4 的第十部分,是由 ITU-T 视频编码专家组(VCEG)和 ISO/IEC 动态图像专家组(MPEG)联合组成的联合视频组(Joint Video Team,JVT)提出的高度压缩数字视频编解码器标准。H.264 编码器和解码器的功能组成分别如图 9-2 和图 9-3 所示。

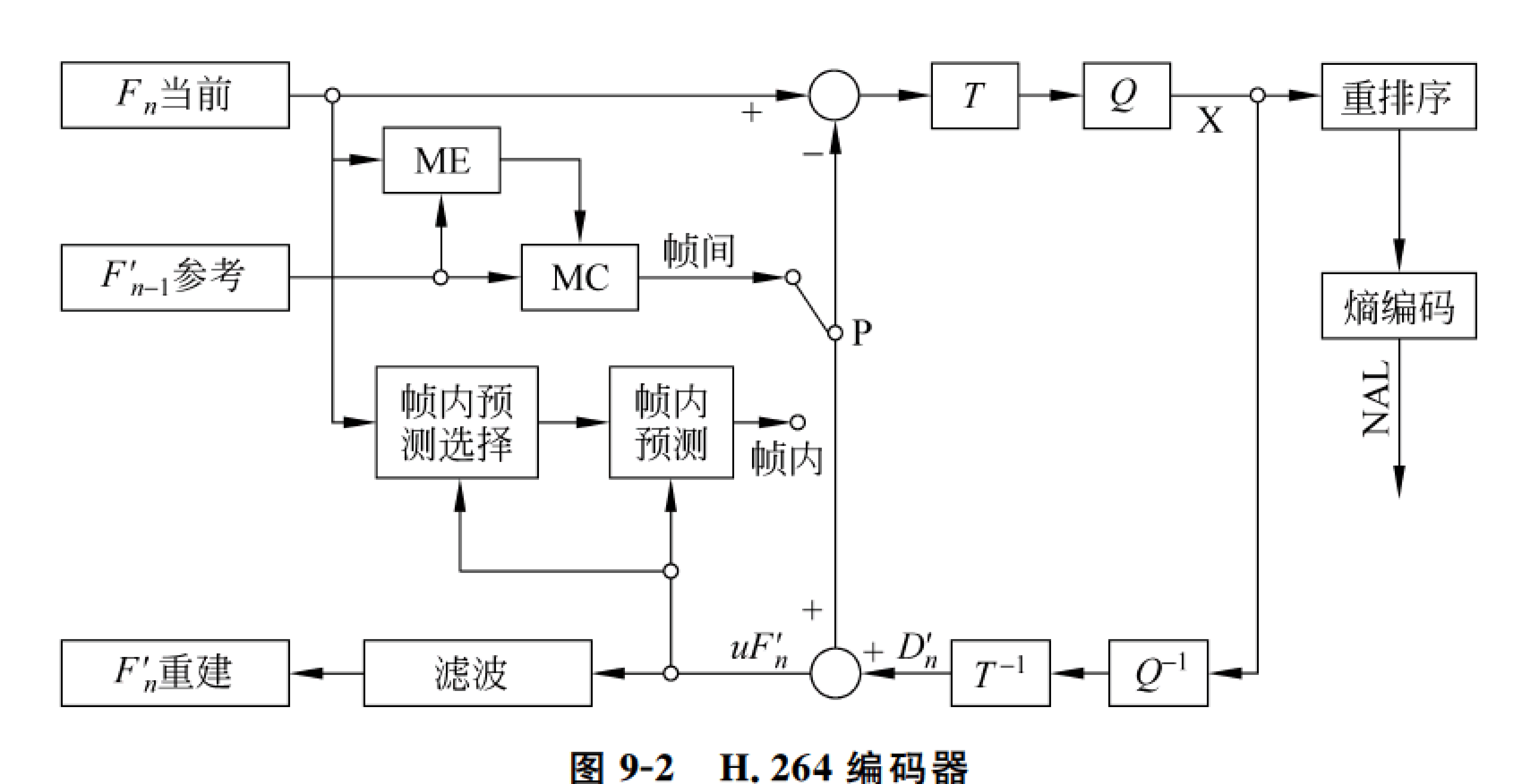


图 9-2 H.264 编码器

H.264 集中了以往标准的优点,在许多领域都得到突破性进展,使得它获得了比以往标准好得多的整体性能。一般来说,H.264 的数据压缩率在 MPEG-2 的 2 倍以上、MPEG-4 的 1.5 倍以上。从理论上来说,在相同画质、相同容量的情况下,可比目前的 DVD 多保存 2 倍以上时间的影像。

H.264 是在 MPEG-4 技术的基础之上建立起来的,其编解码流程主要包括五个部分:帧间和帧内预测(estimation)、变换(transform)和反变换、量化(quantization)和反量化、环路滤波(loop filter)、熵编码(entropy coding)。

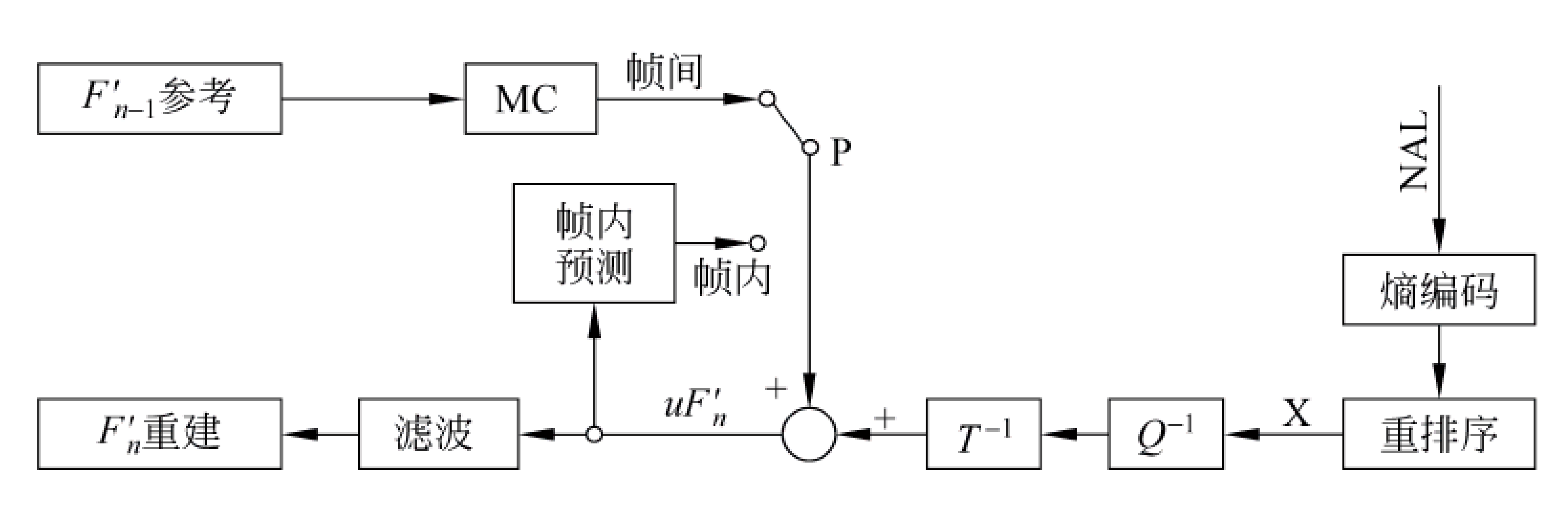


图 9-3　H. 264 解码器

1）帧内预测编码

帧内编码用来缩减图像的空间冗余。为了提高 H. 264 帧内编码的效率，在对一给定宏块编码时，首先可以根据周围的宏块预测（典型的是根据左上角的宏块，因为此宏块已经被编码处理），然后对预测值与实际值的差值进行编码，这样，相对于直接对该帧编码而言，可以大大减小码率。

2）帧间预测编码

帧间预测编码利用连续帧中的时间冗余来进行运动估计和补偿。H. 264 的运动补偿支持以往的视频编码标准中的大部分关键特性，除了支持 P 帧、B 帧外，还支持一种新的流间传送帧——SP 帧。码流中包含 SP 帧后，能在有类似内容但有不同码率的码流之间快速切换，同时支持随机接入和快速回放模式。

3）整数变换

在变换方面，H. 264 使用了基于 4×4 像素块的类似于 DCT 的变换，但使用的是以整数为基础的空间变换，不存在因为取舍而产生误差的问题，此外，整数 DCT 变换还具有减少运算量和复杂度，有利于向定点 DSP 移植的优点。

4）量化

H. 264 中可选 32 种不同的量化步长，步长是以 12.5%的复合率递进的，而不是一个固定常数。且变换系数读出方式也有“之”字形和双扫描两种。

5）熵编码

视频编码处理的最后一步就是熵编码，在 H. 264 中采用了两种不同的熵编码方法：通用可变长编码（UVLC）和基于文本的自适应二进制算术编码（CABAC）。

目前，大多数的视频会议系统均采用 H. 261 或 H. 263 视频编码标准，而 H. 264 的出现，使得在同等速率下，H. 264 能够比 H. 263 减小 50%的码率。也就是说，用户即使是只利用 384kb/s 的带宽，就可以享受 H. 264 下高达 768kb/s 的高质量视频服务。H. 264 不但有助于节省庞大开支，还可以提高资源的使用效率，同时令达到商业质量的视频会议服务拥有更多的潜在客户。

4. AVS

AVS（Audio Video coding Standard）是中国自主制定的音视频编码技术标准。当前，AVS 视频主要面向高清晰度电视、高密度光存储媒体等应用中的视频压缩。

AVS 视频当中具有特征性的核心技术包括 8×8 整数变换、量化、帧内预测、1/4 精

度像素插值、特殊的帧间预测运动补偿、二维熵编码、去块效应环内滤波等。

1）变换量化

AVS的8×8变换与64级量化，可以在16位处理器上无失配地实现，完全适应不同的应用和业务对码率和质量的要求。

2）帧内预测

AVS的帧内预测技术沿袭了MPEG-4 AVC/H.264帧内预测的思路，但AVS亮度和色度帧内预测都是以8×8块为单位的。亮度块采用五种预测模式，色度块采用四种预测模式，而这四种模式中又有三种和亮度块的预测模式相同。在编码质量相当的前提下，AVS采用较少的预测模式，使方案更加简洁，实现的复杂度大为降低。

3）帧间预测

帧间运动补偿编码是混合编码技术框架中最重要的部分之一。AVS标准采用了16×16、16×8、8×16和8×8的块模式进行运动补偿，而去除了MPEG-4 AVC/H.264标准中的8×4、4×8、4×4的块模式，目的是能更好地刻画物体运动，提高运动搜索的准确性。实验表明，对于高分辨率视频，AVS选用的块模式已经能足够精细地表达物体的运动。较少的块模式，能降低运动矢量和块模式传输的开销，从而提高压缩效率、降低编解码实现的复杂度。

4）熵编码

AVS熵编码采用自适应变长编码技术。在AVS熵编码过程中，所有的语法元素和残差数据都是以指数哥伦布码的形式映射成二进制比特流。采用指数哥伦布码的优势在于：一方面，它的硬件复杂度比较低，可以根据闭合公式解析码字，无须查表；另一方面，它可以根据编码元素的概率分布灵活地确定以k阶指数哥伦布码编码，如果k选得恰当，则编码效率可以逼近信息熵。

AVS-视频目前定义了一个档次（profile）：基准档次。该基准档次又分为四个级别（level），分别对应高清晰度与标准清晰度应用。

AVS视频的主要特点是应用目标明确，技术有针对性。因此在高分辨率应用中，其压缩效率明显比现在在数字电视、光存储媒体中常用的MPEG-2视频提高一个层次。在压缩效率相当的前提下，又较MPEG-4 AVC/H.264的主应用模式（main profile）的实现复杂度大为降低。

9.1.5 数字视频内容安全技术分类

随着数字视频技术的发展，基于互联网的视频商业应用越来越普及，也不可避免地出现数字视频内容安全与版权的问题。如何对数字视频内容的安全性进行保护亟待解决。与普通文本数据不同，视频具有数据量大、实时性要求高等特点，为视频安全问题提出更多难题。

当前数字视频所面临的主要安全问题有以下几个方面。

1. 数字视频内容保密

数字视频应用的范围扩展到经济、军事、政治、教育等各行各业。政治、经济、军事等

对其安全性要求很高,因为视频数据本身没有被加密,因此,在传输过程中,很容易被窃取,这需要通过加密等方法来保护这些重要信息。这也是对视频信息的主要保护手段。

2. 数字视频版权保护

为了表明对数字视频作品内容的所有权,通常需要在数字视频作品中含有所有者信息,这样就可以保护所有者的权益。在数字视频作品发行体系中,可以通过一种拷贝保护机制,即不允许未授权的媒体拷贝。在一个封闭或私有的系统中,数字视频需要特殊的硬件来拷贝和观看使用,在视频作品中有标识允许的拷贝数,每拷贝一份,进行拷贝的硬件会修改水印内容,将允许的拷贝数减一,以防止大规模的盗版。

3. 数字视频隐写与分析

敌特机构、恐怖组织等可能将国家政治、经济等机密信息隐藏在视频中进行传递,用于计划和组织破坏活动,使公共信息网络成为破坏社会稳定、危害国家安全和公共安全的通信工具。

4. 数字视频内容认证

目前许多视频编辑和处理软件可以轻易地修改、伪造数字视频的内容,使得视频内容不再可靠。因此需要一种方法进行内容认证和完整性校验来检测数字视频作品,判断其真实性、完整性及原始性。

9.2 数字视频内容加密技术

9.2.1 数字视频加密技术概述

视频加密技术是指一种为了提高视频数据的保密性而对视频数据进行处理的技术。加密以后的数据可以在公开信道上安全地传输而不必担心未授权用户非法获取视频信息。

数字视频在许多方面与静止图像有相同的特性。为了保证流畅的视觉效果,视频加密必须考虑实时在线能力,要求有很高的处理速度。

数字视频的加密算法大致有两类:完全加密算法和选择性加密算法,如图 9-4 所示。

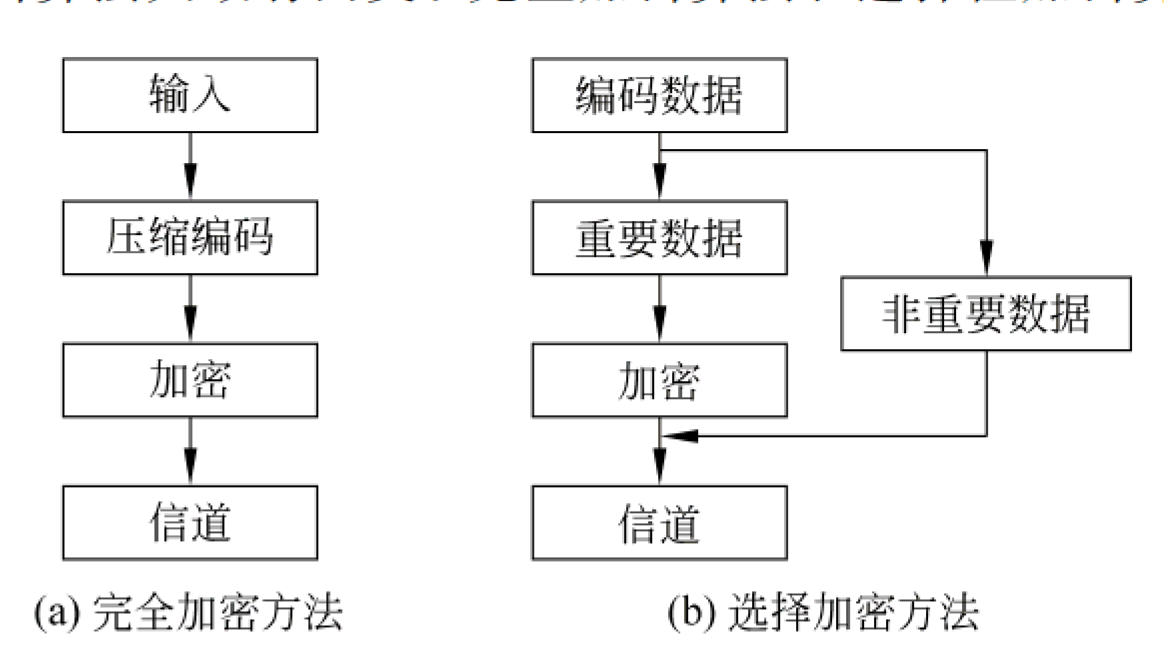


图 9-4 加密方法

1. 视频加密的密码学方法

最早的高密级视频加密方法是对全部视频数据流直接用密码技术加密和解密，国内一般称传统加密方法。

从传统的数字加密技术来看，可以把视频看作文本数据，用经典的加密算法（如DES、AES）进行加密再传输。由于密码技术已有许多安全可靠的成熟算法，以二维或多维数据表示的图像及视频传输和存储时都要映射（如编码）变成一维数据，若不考虑具体数据特征，易于直接应用已有的成熟加密算法。其安全性评价取决于所用密钥（在目前的图像和视频加密方法中安全性最高）。

但从视频的应用特点来看，存在两种困难，一是视频数据量巨大，二是视频在网络中总是以压缩的形式传输。这使得用传统的加密技术直接加密视频数据不可行：由于加密破坏了视频数据空间相关性，使得压缩效率降低，无法达到网络传输所需的码率；经典的加密方法会带来巨大的运算负荷和时延，不能满足视频实时传输播放的要求。另外，标识信息经加密无法识别，就不能实现在线处理，也留下明文攻击隐患。

2. 选择性加密方法

传统加密方法在很多场合难以实用。为此，人们研究视频的信源特征，把密码原理与视频技术结合起来，只对部分重要数据进行加密，这就是选择性加密。

1）选择加密不同帧和块

最基本的选择性加密方法是基于 MPEG 的 IPB 帧结构的，即仅加密 I 帧即可达到加密整个视频序列的目的，因为 I 帧为关键帧，P 帧是在 I 帧的基础上经预测编码得到的，B 帧是在 P 帧和 I 帧的基础上经过内插得到的，所以从概念上讲，如果不知道相应的 I 帧，仅有 P 帧和 B 帧是没有用的。

2）加密运动矢量算法

数字视频采用运动矢量预测技术，减少了编码的数据量，但它包含了大量的原始视频图像中的重要物体信息，所以随机改变运动矢量的符号位或同时改变其数值，再加上运动矢量在预测时的累积效应，几帧之后就有面目全非的效果。

3）DCT 系数置乱

目前绝大多数视频编码标准都采用了 DCT 变换，变换编码是视频压缩中的一种重要压缩手段，在量化后（特别是 zig-zag 排序之后）系数被排在高低频上，通过置乱、改变符号或数值等方法破坏 DCT 系数的原有特性，来获得非常好的加密效果。

4）加密数据格式信息

加密数字视频流的格式信息，如图像序列头信息、宏块的头信息等，以达到非授权者不可同步的目的。对格式信息加密可以明显降低计算量，但也存在缺点：一是头部信息有很多标准信息，攻击者易猜测；二是有很多控制信息，在传输过程中被中间节点用于同步、误码监测等，一旦被加密，中间节点很难处理。

3. 数字视频数据加密的一般要求

视频数据具有数据量大、实时性要求高、冗余度大等特点，压缩后视频数据要求具有数据位置索引等功能，针对数字视频的特殊要求，视频数据加密通常需要满足以下要求：

(1) 安全性。安全性是数字视频加密的首要要求。如果攻破加密算法所花费的代价比视频内容本身的价值还要高时，就可以认为该算法是安全的。

(2) 实时性。在很多视频应用场合，比如视频会议、可视电话等都要求视频传输具有实时性能，因此视频加密算法本身复杂度不能太高。

(3) 压缩比不变性。"加密和压缩天然是矛盾"。在编码过程中设计加密算法，往往会影响视频的压缩效率。视频加密算法应该尽可能地保证压缩比不降低。

(4) 格式符合性。加密前后视频数据的格式没有发生改变，解码器仍然可以读取、播放加密以后的视频，只是读出的视频可能会达不到质量要求或者视频没有意义。

(5) 容错性能。由于视频的传输通道中存在干扰，因此加密算法应该有一定的抗噪能力，有效防止错误扩散。

(6) 数据的可操作性。对加密后的视频数据可直接操作，而不必进行先解密再加密的繁琐过程。这些操作可能包括码率控制、图像块剪贴和增删等。

(7) 安全等级的可控性。通过调节加密参数，可以较好地控制加密强度。

9.2.2 数字视频加密典型算法

到目前为止，国内外的学者和研究人员已经提出了很多种视频加密算法，而且大都是针对 H.26X 和 MPEG-X 两种系列的视频压缩国际标准的选择性加密算法。

相对于完全加密而言，选择加密处理的数据量少、实时性高。因此主流的视频加密方案都倾向于对视频数据进行选择加密。但是选择加密算法也存在问题：各个指标之间相互制约。比如，安全性要求越高，加密的数据量可能越大，加密的速度也就越慢；加密和压缩总是存在着矛盾，在编码过程中设计加密算法，可能会破坏原有数据的分布特性，从而降低压缩效果等。

1. 完全加密

完全加密是一种最直接的加密方法。它没有考虑视频数据的特殊结构，对整个视频流采用标准的加密算法，如用经典的 DES 算法进行加密，又如 VEA 通过与密钥相应位的异或来改变 AC 和 DC 系数的符号位进行了加密。此算法相对其他完全加密算法速度快，适合实时传输，但易受明文攻击，不能提供很高的安全性，只适合代价较低的多媒体应用。

完全加密算法采用的都是传统经典算法(如 DES、AES 等)，它们对穷举攻击、只知密文、已知明文和选择明文攻击都有很好的防范效果；其视频加密系统具有很高的安全性；但是由于视频数据量本身就很大，而且这些传统经典加密算法也都具有很高的计算复杂度，因此，完全加密算法的计算复杂度较高。而且完全加密算法将视频数据当作一般的二进制数据进行加密，没有考虑视频数据的数据格式，不具有数据可操作性。

2. 选择性加密

1）空域选择加密

空域选择加密是指在空域中设置加密算法，如对空域数据进行置乱、替换或者加密运动矢量等。Spanos 和 Maples 提出的 Aegis 安全系统属于空域选择加密。该系统采用 DES 等算法对 I 帧进行加密，此外，它还加密了视频序列的头部信息。

为了进一步隐藏 MPEG 的标识信息，Aegis 系统对 MPEG 视频流中的 32 位 ISO 码也做加密，以进一步提高安全性。该算法运算量小、实时性好，但是安全等级不够。在 P 帧和 B 帧中残留有未经预测的 I 块，这样通过预测累加可以恢复出部分视频内容。而且对头信息的加密将使算法不具有格式符合性。

2）变换域选择加密

Tang 提出的“之”字扫描扰乱重排法，是将 64 个 DCT 系统完全置乱，以此实现加密。但这违背了“之”字扫描的能量大小排列顺序，降低了压缩效率。根据 L. Qiao 等分析表明，采用随机置乱方式代替“之”字扫描，不但大大降低了压缩比，而且使得密码系统不能抵抗已知明文攻击。

A. S. Tosum 等对 Tang 的算法做了改进，将 64 个 DCT 系数按照频带划分为三段。Tosum 建议的分层方法为(4，19)，即第一层为 1～4 点，第二层为 5～18 点，第三层为 19～64 点。在每一层内置乱相对于 64 点完全置乱，能够获得较高的压缩比，但这是以降低安全性为代价的。

Shi 用异或运算只改变帧内宏块的 DCT 系数符号位和运动矢量符号位，基本思想是用随机产生的密钥流与 DCT 系数符号进行按位异或运算，然后将加密后的符号相应地赋回给原数据。该算法大大降低了运算量，速度快，能够满足实时性的要求，但不能提供可靠的安全性，攻击者可以假定各系数的符号均为正(或负)，这样即可获取部分信息。

3）熵编码过程加密

熵编码过程加密是指在熵编码的过程中设计加密算法。这类加密算法一般具有较高的安全等级和较小的数据膨胀，但是通常算法比较复杂，实现起来有一定难度。Jiangtao Wen 等提出的码字序号加密法就是其中一种，它将加密区域分为定长编码和可变长编码，给码表的每一个码字分配一个索引序号，对串联索引序号进行加密，而不是对码字内容进行加密，然后将加密的索引序号映射回原来的码表中，由此完成码流的加密。

该算法具有很好的格式符合性和安全性，但是会涉及拆分 VLC 码表，算法复杂，不易实现。

4）部分码流加密

为了减少加密的数据量，部分码流加密选择了对压缩流中的部分码流进行加密。例如，可以将编码以后的码流分层若干个“码段”，然后选择部分“码段”进行加密。由 Qiao 和 Nahrstedt 提出的 VEA 加密算法可以归为部分码流加密。该算法首先将需要加密的码流分成 128 个字节的块，再根据密钥将每个块分成奇偶两列，然后进行异或运算。接下来将异或的结果与先前奇偶列中的一列(此列经过了 DES 加密处理)级联输出。经过这样的处理，运算量大大降低。Ali Saman Tosum 对 VEA 进行了改进，对序列进行两次

奇偶分离。这样加密的数据量又减少了一半。VEA 加密算法的优点是在获得较高安全等级的同时大大降低了运算量，并且保持了压缩比不降低；缺点是不具有格式符合性、不具备安全等级可控性和数据的可操作性。

选择性加密算法仅仅对整个压缩编码过程的某一个或某几个模块进行加密，所以其加密效果和抗攻击性都不如完全加密算法。由于选择性加密算法在计算复杂度、视频传输实时性方面带来的好处远大于其在安全性上的不足，所以目前越来越多的科研人员把研究重点放在了选择性加密上，如何增强其安全性成为最重要的因素。

3. 基于混沌的视频加密技术

混沌序列是一种非线性序列，其结构复杂，难以分析和预测，混沌系统可以提供具有良好随机性、相关性和复杂性的伪随机序列，这些都是很有吸引力的特性，使其有可能成为一种可实际被选用的流密码体制。从英国数学家 Mattews 明确提出用混沌系统来产生序列密码以及后来 Pecora 和 Carroll 提出混沌自同步方法以来，混沌同步保密通信在国际电子工程界得到了广泛的研究。

选用何种混沌系统能产生满足密码学各项要求的混沌序列是目前各国密码学者大力研究的问题。比较典型的有 1989 年，Matthews 提出用 Logistic 混沌映射改进成的迭代混沌系统。1992 年，Pecora 和 Carroll 提出著名的 Lorenz 系统。法国 Besancon 大学 Goedgebuer 等利用可调激光二极管研制了一个光传输数据的系统。

Shannon 在信息论中证明，要实现完全的保密，即具有完全的抗破译能力，必须能够产生无限长随机序列密码，此所谓“一次一密”。但是在设计密码系统时要产生比明文序列还要长的随机序列非常困难和复杂度高并且也会给密钥管理带来极大的不便。因此在实际应用中往往使用随机性能略低于随机序列的伪随机序列作为代替，进行加密。显然，高性能的伪随机密钥流的产生和同步是决定系统性能的关键因素。

时空混沌流加密的算法简单，只需将产生的混沌伪随机二进制序列与编码产生的视频压缩流逐位进行位操作即可。由于时空混沌序列的高度伪随机特性，因此该加密算法具有很高的安全等级。

9.3 数字视频隐写与水印技术

9.3.1 数字视频隐写技术

目前的隐写技术大部分还是集中在静止图像方法，并出现了一些比较成熟的方法，如空间域的 LSB 替换隐写、相邻像素对差分隐写等。虽然数字视频可看作一系列静止图像组成，但如果直接把针对静止图像的隐写方法用于数字视频，则会存在不能准确提取或影响视频质量等问题，再加上具体视频编码压缩格式的限制，更容易造成技术应用上的局限性。因此，数字视频隐写技术需要根据视频数据的特点来设计合适的隐写算法。

1. 数字视频隐写的通信模型

Simmons 于 1983 年提出的“囚犯问题”是一个经典的隐写系统，根据数字隐写的囚犯模型，视频数字隐写的典型通信模型如图 9-5 所示。在该通信模型中，发送方使用一定的信息嵌入方法和密钥将秘密信息 m 嵌入到载体视频 c 中，形成载密视频 s。在公共信道中传输时，s 可能会遭受各种处理与攻击。接收方得到经过各种处理与攻击的载密视频 s'后，按照信息提取方法和双方共享的密钥从中提取出秘密信息 m'，从而完成隐蔽通信过程。

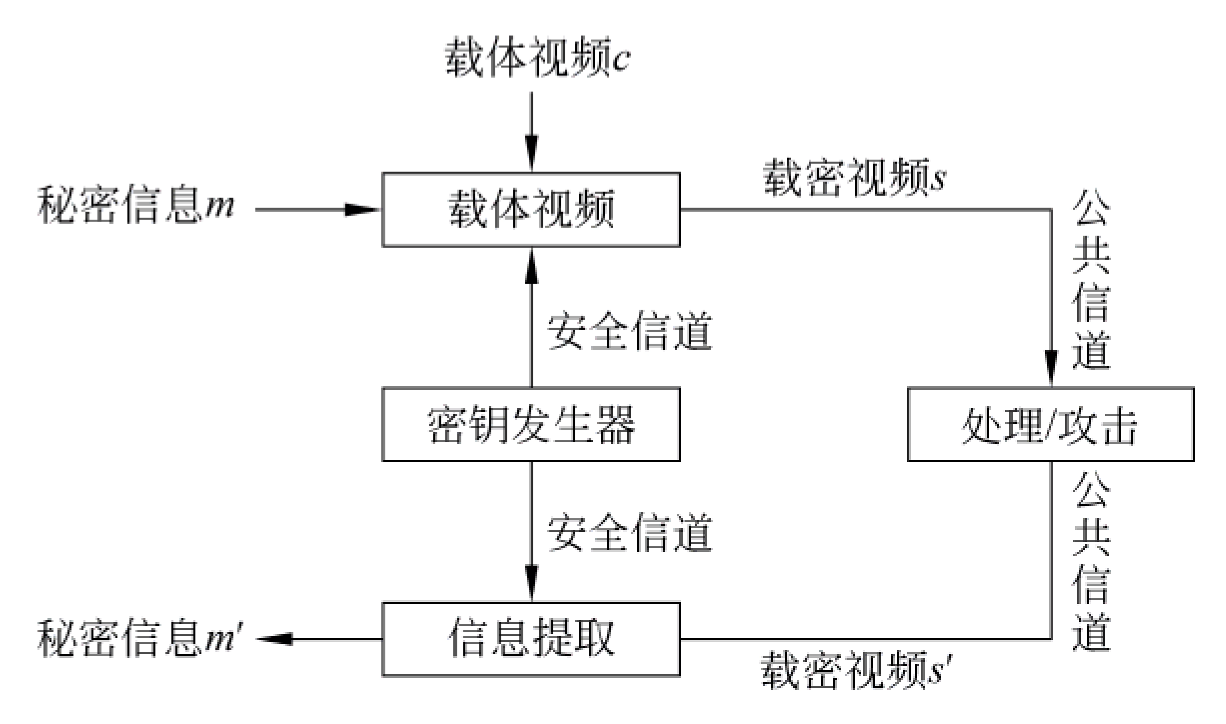


图 9-5 数字视频隐写的通信模型

为了对抗处理，需要考虑的是鲁棒性问题，而对抗攻击时考虑的则是安全性问题。

2. 数字视频隐写的特点

因为视频序列基本上可以看成是一系列连续的静止图像组，所以它在隐藏技术的应用模式和设计方案上与静止图像是非常类似的。但因其具有更大的可用载体空间，在时间域上具有特殊的压缩特性，以及视频应用系统本身对实时性等的约束，都导致了视频信息隐藏技术具有自身的特殊性：

(1) 视频信息隐藏具有更大的可用载体空间。

(2) 视频信息隐藏系统必然要经历有损压缩编码过程，数字视频隐写不仅要考虑空间域和变换域统计特性的影响，还要考虑相邻帧间统计特性即时间域统计特性的变化。

(3) 具体的视频编解码应用系统对信息隐藏和提取算法的实现提出了实时性或准实时性的要求和其他一些约束条件(如恒定码率)。

由于视频信息隐藏区别于静止图像信息隐藏的这些特殊性质，现有的图像隐藏算法还不能很好地保护视频数据，视频信息隐藏技术面临着新的挑战。

3. 数字视频隐写技术分类

根据隐秘信息嵌入视频中的时机不同，视频中的数据嵌入策略可分为以下三类(嵌入策略分类位置如图 9-6 所示)。

(1) 将数据信息直接嵌入到原始视频图像中，形成含隐藏数据的视频信息后，再进行视频编码。已有大量的数据嵌入法都是基于此方案来进行的，如时空域水印算法。在视频数据编码前或完全解码后的静态序列图中，利用人眼视觉感知模型 HVS 或其他基于

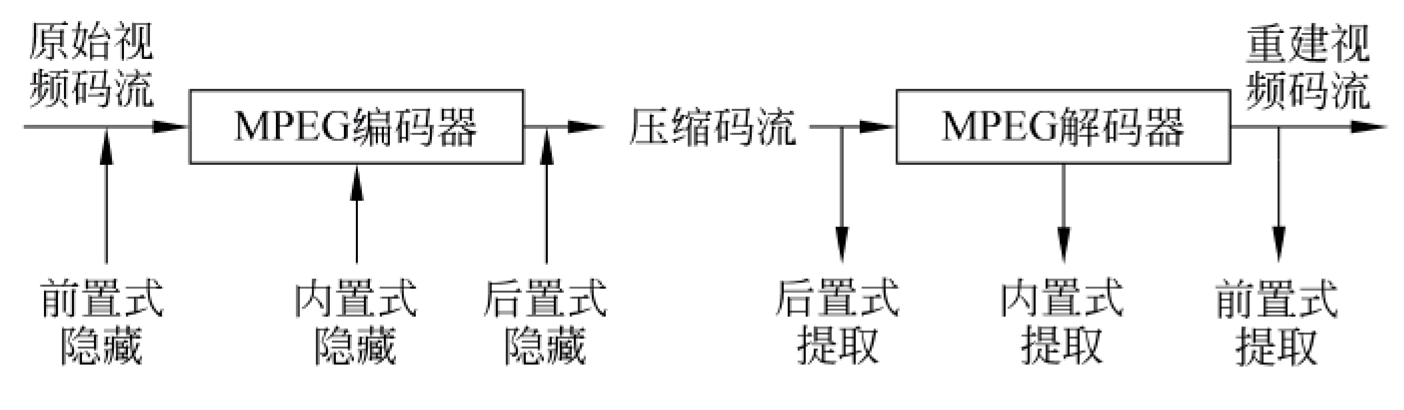


图 9-6 嵌入策略分类位置图

纹理、区域等技术，使用空域、变换域隐藏法，将数据嵌入到图像中，再将图像进行编码形成带有隐藏数据的视频流。

(2) 将嵌入过程引入到视频编码器中。这一方案虽然增加了嵌入算法的局限性，但是由于其通常是通过调制 DCT 变换或量化之后的系数来完成信息嵌入过程的，这样可以通过自适应的机制依据人的视觉特性进行调制，在得到较好的主观视觉质量的同时又能获得较强的鲁棒性。

(3) 把水印信息直接嵌入到视频压缩码流中，其最大的优点在于不需要完全解码和再编码过程，降低隐藏算法的复杂性，因此对整体视频信号造成的影响较小。但数据的嵌入必须满足视频系统对视频压缩码率的约束，同时嵌入过程可能造成对视频解码系统中运动补偿环路的不良影响，因此该类算法设计具有一定的复杂度。

9.3.2 数字视频水印技术

数字视频水印就是加载在数字视频上的数字水印，它利用视频数据中普遍存在的冗余数据和随机性把表征版权的信息嵌入到原始视频数据中，从而保护数字产品的版权或完整性，确保版权所有者的合法权益。

1. 数字视频水印的基本特征

视频数字水印首先应该具有数字水印的一般特征，如安全性、可靠性、鲁棒性、不可感知性。对于视频水印而言，由于相邻帧之间内容的高度相关性，以及可能遭受的如帧丢弃等各种攻击，数字视频水印还有一些特殊的要求，例如：

(1) 实时处理性：数字视频水印的嵌入和提取应该具有低复杂度，必须在短时间内完成，以保证视频数据的实时编解码。

(2) 随机检测性：可以在视频的任何位置、在短时间内(不超过几秒钟)检测出水印。随机检测性比实时性具有更严格的要求：一个水印方案是实时的，但是如果只能从视频的开始位置按播放顺序一步步检测出水印，则不具有随机检测性；如果跳转到视频的任何一个位置，也能够在很短时间内检测出水印，则具有随机检测性。

(3) 与视频编码标准相结合：视频数据由于其数据量极大，在存储、传播中通常先要对其进行压缩，现在最常用的视频数据压缩编码标准是 MPEG-4 和 H.264。如果是在压缩视频中嵌入水印，很显然是与视频的压缩编码标准相结合的；如果是在原始视频中嵌入水印，由于水印嵌入是利用视频的冗余数据来携带信息，而视频压缩编码则需要除去视频中的冗余数据，如果不考虑视频压缩编码标准而盲目地嵌入水印，则嵌入的水印很

可能在编码过程中就完全丢失了。

(4) 盲水印方案：水印检测时不使用原始视频数据，以确保水印检测能够实时完成。

(5) 鲁棒性：对于视频数据，必须保证水印方案对一些无意或故意的处理和攻击的鲁棒性，如帧平均、帧丢弃等。在视频上的任何处理，只要没有将视频破坏到失去使用价值的地步，都应该不会破坏所嵌入的水印。

(6) 视频速率的恒定性：水印嵌入视频数据后不能改变视频流的码率，必须服从传输信道规定的带宽限制，否则将有可能造成解码后的视频图像和声音的失步，降低视频的质量。

2. 数字视频水印技术分类

对于视频水印技术，可作如下分类：

(1) 按载体类型分类：包括基于原始视频的方法和基于压缩视频的方法。基于原始视频的水印算法，是对未经编码的视频流数据直接进行处理，在原始视频数据中嵌入水印。基于压缩视频的水印算法，则与某种视频压缩标准，如常见的 MPEG-1、MPEG-2 或 MPEG-4 相结合，在编码视频数据中嵌入水印。

(2) 按嵌入域分类：主要可分为空域(或时域)方法及变换域(频率域)方法。空域替换方法是用待嵌入的信息替换载体信息的冗余部分。一种简单的替换方法就是用待嵌入消息位替换载体中的一些最低有效位，只有知道隐藏信息嵌入的位置才能提取信息。变换域方法是在宿主信号的某个变换域，如 DCT 或小波域中嵌入信息。

(3) 按密钥分类：若嵌入和提取采用相同密钥，则称其为对称水印，否则称为非对称水印，也称为公钥水印。

(4) 按检测时是否需要原始宿主信号分类：分为盲水印方案和非盲水印方案。正如前面所讨论的，一般来说，视频水印方案在检测时通常不需要原始的宿主信号。但是，也有极少数方案需要原始的宿主信号。

(5) 按水印特性分类：可以将数字水印分为鲁棒水印和脆弱水印两类。鲁棒水印能够经受各种有意或无意的攻击；脆弱水印则对于信号的改动比较敏感，主要可用于篡改提示。这里主要是对鲁棒水印进行介绍。

(6) 按水印的可见性分类：分为可见性水印和不可见性水印。现在一般研究的是不可见的水印，但是在一些应用中可能需要嵌入可见的水印，比如在视频中嵌入标识信息，对拷贝行为提出警告。

另外还可以按用途分类、按内容分类等。

9.3.3 数字视频隐写与水印典型算法

1. 数字视频隐写技术典型算法

1) 前置式隐藏技术研究

前置式隐藏法(如图 9-6 所示)最大的优点就是可以借鉴已存在的成熟图像隐藏方法来进行嵌入操作。如可通过将水印或隐藏图像先进行量化，后使用多维网格进行编码，

再根据嵌入图的纹理特征，将编码后隐藏信息嵌入到载体图的 DCT 域中。要嵌入一 8×8 DCT 的隐藏图像块只需要 16 个载体视频的 DCT 块。这样不仅可以抵抗 MPEG 压缩，还能获得比较高的信息隐藏量。

Swanson 提出了一种多尺度水印隐藏法，利用小波变换及 HVS 视觉模型，将数据信息藏入到视频静态帧中，来达到隐藏效果。此方法鲁棒性极高，可抵抗噪声干扰、MPEG 压缩，甚至帧重组等情况。但缺点是复杂性太高，对每一帧隐藏数据都同时引入了小波变换和 HVS 模型。

这些方案虽利用了比较成熟的数字图像隐藏策略，但因其未考虑视频特性，在经过编解码处理后，势必造成部分数据的丢失，为数据的恢复和提取带来很大的不利因素。其缺点如下：

(1) 会增加视频码流的数据比特率。

(2) 经 MPEG 压缩后会丢失水印。

(3) 容易降低视频质量。

(4) 对于已压缩的视频，需先进行解码，嵌入水印后再重新编码。这样带来的系统计算量较大，无法满足许多应用系统的要求。

2) 内置式隐藏技术研究

目前有很多视频隐藏法都是基于此种方法来进行研究的，既考虑了视频压缩与静止图像压缩相似的频域变换特性，又利用到视频文件本身在速率和时间域上的特点。如将视觉分析和块分类技术应用到视频隐藏中动态选取 DCT 频域中的最佳数据嵌入系数。它将 DCT 亮度块按其能量分布分为低活动、边缘、垂直边缘、水平边缘和纹理五大类，并根据块类型的不同分别将数据隐藏到不同的 AC 系数中。

Linnartz 提出了另一种基于 MPEG-2 中 GOP 图像组的压缩域嵌入法，利用 GOP 结构来嵌入数据，每个 GOP 图像组中可嵌入 6 比特数据。此方法只适用于压缩过程进行，不能在 GOP 结构固定后进行操作，因此也就无法应用于需使用比率调整来优化编码效率的系统，因为这些系统通常不限制在预定的 GOP 中。基于上述原因，该方法在抵抗解压和重压缩操作的鲁棒性方面性能较弱。

由于视频 I 帧上色度的 DCT 直流系数 DC 是一个始终在视频流中存在、且鲁棒性很强的系数，所以戴元军提出，将数字信息经序列调制后加入到 I 帧的色度 DCT 直流系数中，这样嵌入信息的鲁棒性就会较高。同时，为了保证视觉上的不可见性，在修改 DC 值的同时，要求其干扰低于一定的门限值，根据系数的大小自适应地加入不同强度的水印，从而实现低隐藏量但高鲁棒性的视频隐藏效果。这一类方案的优点是：

(1) 数据仅嵌入在 DCT 系数中，不会增加视频流的数据比特率。

(2) 易设计出抗多种攻击的数据嵌入算法。

(3) 适用于所有基于 DCT 变换的视频编码。

缺点是在操作不当时容易引起视频质量的下降。

3) 后置式隐藏技术研究

在后置式隐藏技术类算法中，信息隐藏技术已逐渐把时间域方面特性的利用放到研究的主要层面上来，在追求简单和实时应用方面起到了很不错的效果。H&G 算法作为

该类算法中的一个典型模型，其基本设计流程如图 9-7 所示。

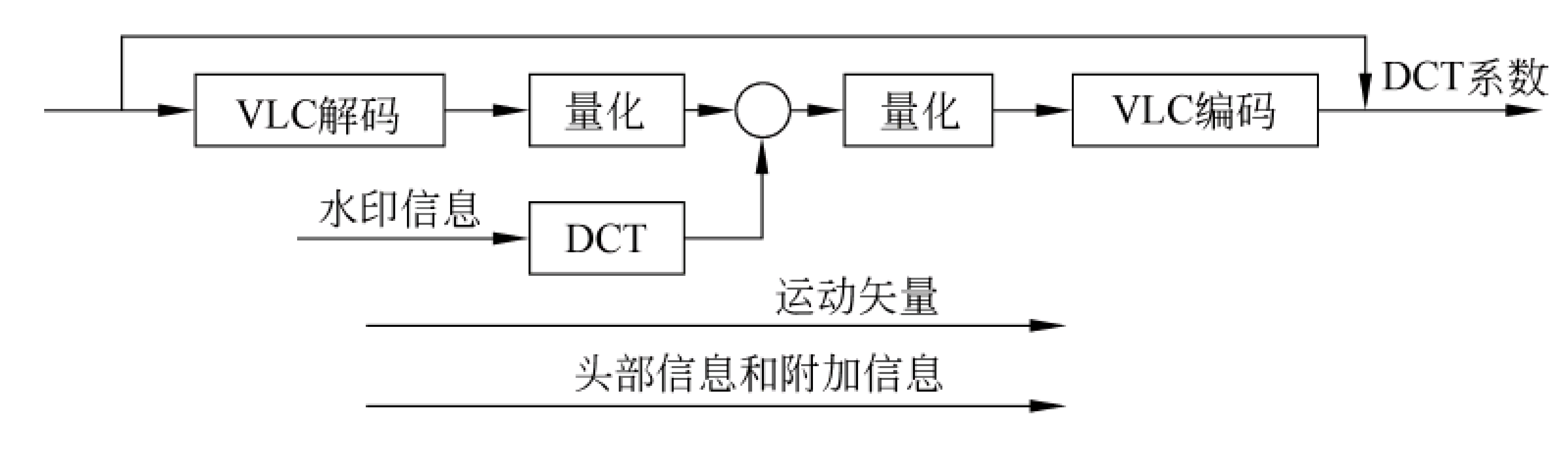


图 9-7　H&G 算法流程图

H&G 算法首先将水印扩频与自适应调整成为与视频序列相同尺寸的空间域水印图像，再作 8×8 DCT 变换，然后利用频域的线性特性，将变换系数与重建视频 DCT 系数块进行线性叠加来完成数据嵌入操作。为避免修改数据造成的视觉影响，还同时引入了飘移补偿方案来提高重建视频图像质量和系统稳定性。

Jordan 提出将数据藏入压缩视频的运动补偿中，因为运动补偿不易产生视觉影响，且嵌入信息可不需解码就直接可从运动向量中提取，即使视频解码后也可通过一步压缩来重提数据。此算法复杂性低，但鲁棒性不高，且需统计量足够多才能嵌入数据，数据嵌入量较小。

Langelear 提出了通过修改 DCT 系数的 VLC 编码来将数据直接藏入视频码流中的方法。该算法的思想是通过在比特流中找到一个相似存在的 VLC，将其进行替换，使嵌入比特能藏入到其中。

朱仲杰提出了将水印直接藏入视频码流的运动矢量上。为达到水印不可见的效果，每个画面组（一般 12 帧图像）中只藏入一幅水印图像，将数据嵌入到所有的 B 帧和 P 帧中。并通过在计算划分好的运动矢量组中随机选取某一嵌入位置，根据运动矢量特征值与要嵌入数据的关系，来简单修改运动矢量，使 P 值变换为 1 和 2 来表示为 0 和 1 的嵌入，从而达到隐藏的效果。其中，V_x 和 V_y 分别表示选取运动矢量的水平和垂直分量。该算法简单、有效，并实现了数据的盲提取功能。

由于这类方案将数据直接嵌入到 MPEG 压缩比特流中，其显著的优点就是没有解码和再编码的过程，因而不容易造成视频质量的下降，同时计算复杂度较低。缺点是数据的嵌入必须满足视频系统对视频压缩码率的约束，同时嵌入过程可能造成对视频解码系统中运动补偿环路的不良影响，因此该类算法设计具有一定的复杂度。

2. 数字视频水印典型算法

按照载体类型的不同，数字视频水印可分为基于原始视频的水印方法和基于压缩视频的水印方法，如图 9-8 所示。

1）空域水印

Hartung 等提出了借鉴扩频通信的基本思想在未压缩视频中嵌入数字水印的方法。水印嵌入时，按照空间上的从左到右、从上到下以及时间上的先后顺序，将视频信号看成一个一维信

- 视频水印
  - 原始视频水印
    - 空域水印
    - 频率水印
  - 压缩视频水印
    - 嵌入DCT系数
    - 嵌入运动矢量
    - 嵌入脸部运动参数
    - 嵌入VOP图像类型

图 9-8　数字视频水印的一种分类方法

号：水印信号则经过扩展、放大和调制，得到一个拟随机序列，采用普通的加法将该随机序列加到一维视频信号中，就得到了嵌入了水印的视频信号。

Kalker 等将视频看成一系列的静态图像，在数个连续的帧中嵌入相同的水印。这里利用了扩频的基本思想，水印是一个加性噪声。水印嵌入时，为了在图像活动较多和较少的区域（即纹理较多和较少的区域）采用不同的嵌入强度，并可以采用局部缩放因子。

2）频域水印

Degumaume 等在视频序列的三维 DFT 域中嵌入水印。因为在整个视频序列上进行三维 DFT 耗费巨大，所以首先将视频序列划分为连续的、非重叠的、长度固定的帧序列，水印嵌入或提取分别在每个序列上重复进行，在每个序列中嵌入相同的信息。水印嵌入时，将水印信号编码成扩频信号，对帧序列进行三维 DFT 变换，然后，选择 DFT 系数的中频部分来嵌入水印。该水印方案对于空间位移和时间位移具有固有的不变性，同时，由于扩频序列的特性，该水印方案也能抵御简单过滤、添加噪声、MPEG 压缩等处理。

Swallson 等提出采用三维小波变换的水印方案。小波变换是用多个分辨率表示信号的一个有力的工具。小波分解的多分辨率特性在时域、空域或频域提供了信号的局部特定信息，可以用于信号的分析和处理。基于小波技术的数字水印方案是近几年来的一个研究热点。

3）嵌入 DCT 系数

Hartung 等提出利用扩频的思想在 MPEG-2 压缩视频中嵌入水印的方法。水印信号经过扩展、放大和调制，得到一个伪随机序列，然后对其进行 8×8 的 DCT 变换，并将 DCT 系数叠加到 MPEG-2 码流的 8×8 的 DCT 系数上。这里主要需要考虑两个问题：

（1）由于 MPEG-2 的 DCT 系数是用变长码进行编码的，系数在添加水印前后的编码长度会发生变化，因此，如果要求不增加视频码流的长度，那么，在出现添加水印后 DCT 系数的编码比特数增加的情况时，仍保留原有的系数。

（2）在 MPEG-2 编码方式中，帧间编码帧（P 帧和 B 帧）是从其他帧预测得到的，用一个运动补偿向量来从其他帧重建当前帧。P 帧本身也可能作为其他帧的预测参考，一个帧内的微小变化，会在时间、空间上传播开来。因此，在水印信号之外，需要添加一个偏移补偿信号，来补偿前一帧的水印信号。

Busch、Hsu、Dittmann 等都分别提出了基于 DCT 系数的视频水印方法，这些方法中有部分借鉴了 Koch 和 Zhao 的静态图像水印算法，同时考虑了人类视觉系统的特性，使嵌入的水印满足不可感知性。

4）嵌入运动向量

Jordan 等在一份 MPEG-4 提案中提出了一种直接针对 MPEG-4 编码视频流的水印方法，通过修改运动向量来嵌入信息。

5）嵌入脸部运动参数

Hartung 等还提出在 MPEG-4 脸部运动参数中嵌入数字水印的方法，其中仍然采用了扩频的思想。在 MPEG-4 中定义了一个一般的脸部，并能够通过脸部运动参数（Fadal Animation Parameter，FAP）运动起来，FAP 总共有 66 个。MPEG-4 编码过程中从视频

序列中确定 FAP,可以将 FAP 看成是随着时间变化的一维向量。

6) 嵌入 VOP 图像类型

Linnartz 等提出在 MPEG 编码过程,根据水印信息选择编码视频帧所谓的图像类型(picture type)。其基本思想如下:在 MPEG 中,图像类型分为 I 帧、B 帧和 P 帧。从一个 I 帧开始,直到但不包括下一 I 帧的一系列帧称为一个图像组 GOP,如果将每个 GOP 的长度固定为 12,并且用 B 帧表示比特 I,P 帧表示比特 0,则每个 GOP 和一个二元序列存在一个一一对应关系。将二元序列编码为 Hamming 码,并排除掉一些不常见的序列,可以得到一个有 62 个码字的码表,也就是说,每个 GOP 可以携带近六比特的信息,这对于一些类似嵌入版权所有者信息的应用是足够了的。

9.4 数字视频隐写分析技术

隐写分析研究近几年迅速兴起的原因主要是由于信息隐藏技术的实际应用中可能涉及隐蔽通信这一敏感问题。但大多数研究成果仍然仅限于静止图像的隐写分析,并且集中于较为简单的 LSB 模式的隐藏算法的分析。

目前,针对视频信息隐藏的分析技术发展相对缓慢,一方面是由于视频信息隐藏及其分析技术需要具备视频编解码系统的研究背景;另一方面是目前只有很少成熟的视频信息隐藏软件被公开。但数字视频作为未来网络信息资源的重要组成,基于视频资源的信息隐藏及其隐藏分析技术正逐步成为信息隐藏领域的研究重点。

9.4.1 数字视频隐写分析概述

1. 数字视频隐写分析技术的特点

由于数字视频和静止图像的紧密关联,因此数字视频隐藏分析和静止图像隐写分析在应用模式和设计方案上都具有一定的相似性,一些曾成功应用于静止图像的隐写分析技术可以被直接引入到视频隐写分析之中。但是由于数字视频和静止图像间的差异以及数字视频隐写分析必须与具体的应用系统相结合的特点,使得数字视频隐写分析与静止图像隐写分析相比,它还有自身的特殊性。

(1) 绝对大的隐藏容量和相对小的嵌入比率。由于视频资源自身的数据量要远远大于一幅静止图像的数据量,通常它所体现出来的绝对隐藏容量也很大。一个 900MB 的 DVD 文件可以隐藏约 10MB 的信息,而一幅 512×512 的图像其信息隐藏量往往只有几千字节。但实际上,这种绝对大的隐藏容量往往使人忽略其相对小的嵌入比率,按照上面的例子,静止图像隐藏嵌入算法可以达到总数据量的 10%的嵌入率,而视频信息隐藏的嵌入率最多只有 1%左右。

(2) 对视频编解码系统的强依赖性。较为成熟的视频信息隐藏算法往往对视频编码系统具有较高的依赖性,甚至完全融入编解码系统之中。这是由于视频资源必须经过有损压缩编码系统,并会造成部分信息的损失。如果隐藏算法游离于这些视频编码系统之外,那么视频压缩编码系统就成为这些隐藏系统必须能够抗击的一种特殊攻击模式。例

如一些将视频序列作为一幅幅静止图像进行隐藏处理的空间域信息隐藏技术，在这种高压缩编码条件下，往往会产生大量的检测错误，必须采用扩频等手段来提高隐藏信息检测的正确率，这将是以大幅度缩减实际数据隐藏量为代价的。

(3) 序列图像时间域相关特性的利用。在静止图像隐藏分析算法中，多数是利用图像空间域和变换域的相关特性进行统计分析，而视频信息隐藏系统往往因为高压缩算法的引入使得这些相关特性消失殆尽，但视频系统又提供了时间域的相关特性，而一般单向的压缩编码流程使得隐藏算法很难估计时间域特性的变化，这为信息隐藏分析提供了一个有力的工具。

2. 数字视频隐写分析的设计策略

现行的视频压缩编码系统是通过一套完整的混合体系将多种压缩编码技术整合在一起，和模块间相互联系制约。同时，标准还通过严格的码流语法将编码后的数据流进行格式规范。这些约束为数字视频隐写技术的引入带来许多障碍，但为数字视频隐写分析提供了相应的思路和方法。

从视频码流语法角度看，抛开算法细节，任何信息隐藏算法最终必然是调整视频码流数据中的某些信息，或改变其数值，或改变其位置，使其对嵌入信息进行必要的调制，从而实现特殊隐藏信息的传递。隐藏分析系统如果能够准确地检测出这种数据上的变化，就能够追踪到隐藏信息的藏身之地。

从视频系统结构的角度来看，各种隐藏算法必然要与视频压缩编解码系统相结合起来，并从中选择适当的模块引入嵌入信息，这样就可以使隐藏的信息能够更为有效地融于视频数据之中，同时也可以避免视频有损压缩编码所带来的信息丢失。相应地，隐藏分析系统可以通过分析系统模块之间的相关程度，来定性判断每个模块引入信息隐藏的可能性，并根据局部数据的相关特性定量分析检测出可疑数据。

从隐藏分析角度而言，对于 MPEG-2 压缩码流大致可以分为头部信息、DCT 系数信息、运动矢量信息等几个部分，只要这些数据元素具备随机性和可控性，就可以作为信息隐藏的载体。通过分析视频压缩编码系统以及对视频信息隐藏算法的总结，在视频码流中占有很大比例的 DCT 系数及其对应形成的 VLC 码字、运动矢量等数据元素，同时具备了可控性与随机性，它们在视频压缩编码系统中可能形成的嵌入位置如图 9-9 所示。

嵌入点 1 的信息隐藏算法多来自较为成熟的静止图像隐藏技术，根据其嵌入过程大体可以分为两类：一种是在空间域直接完成信息的嵌入过程；另一种是图像经变换后在变换域嵌入信息，然后再反变换回到空间域。综合分析这类算法可以看出，尽管可由嵌入强度自适应控制机制来保证隐藏信息达到感知不易察觉，但是由于其后的有损压缩编码具有一定的不可知性(例如码率约束、质量约束等)会对隐藏信息带来部分损失，要保证信息的准确传递就必须加大隐藏数据的冗余或者嵌入强度。隐藏分析检测效率会随着嵌入强度的增大而提高。隐藏检测分析的另一个突破口则是这类隐藏算法嵌入信息的构成模式：一种是各帧嵌入相同的信息，另一种则是各帧嵌入不同的信息。

嵌入点 2 的基本算法与嵌入点 1 相类似，只是能够利用编码器中的部分信息对嵌入比特进行调控，这样虽然可以一定程度上提高嵌入调制的自适应能力，但是增加了系统

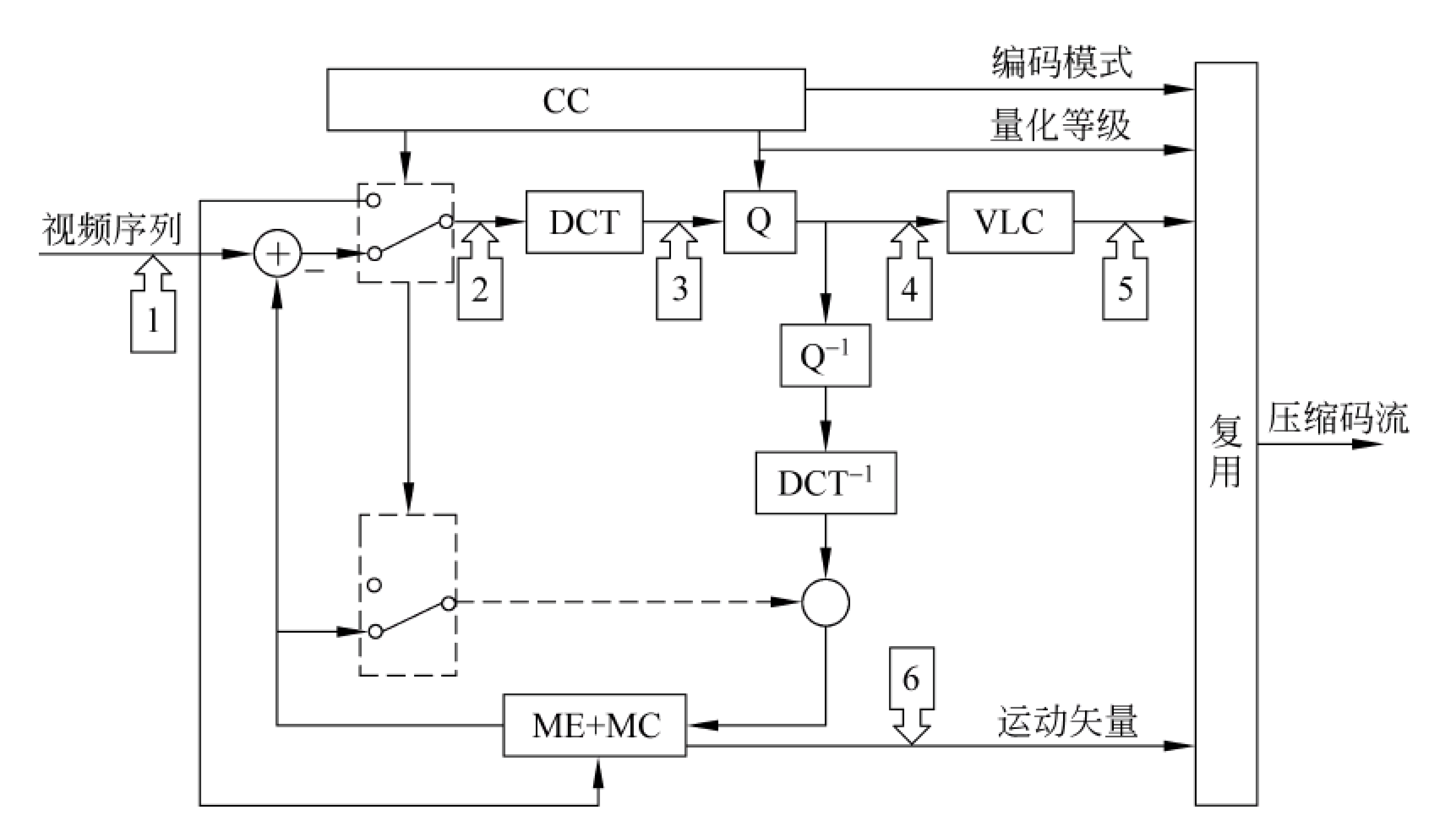


图 9-9 MPEG-4 编码系统中相应嵌入点示意图

复杂度,总体上优势不大,很少被人采用。

嵌入点 3 是在 DCT 域实现信息的嵌入调制的,这便于利用 DCT 域中人眼的频率掩蔽效应实现信息嵌入。

嵌入点 4 选择量化后 DCT 数据作为信息嵌入的载体,这类算法易产生大容量的信息嵌入,然而嵌入点选择在量化之后,使得嵌入调制对图像质量影响较大。

嵌入点 5 是在变长编码后嵌入,可以利用现有压缩码流作为载体,只需对压缩码流进行部分解码,即可实现信息嵌入,但是算法必须考虑编码系统码流格式和传输条件等因素的约束,因而算法设计具有较大难度。

嵌入点 6 选择运动矢量作为信息嵌入点,往往不会直接体现在当前帧的编码质量之中。

从以上分析可以看出,同时具有随机性与可控性的数据可以产生多种多样的嵌入方式,因而对于它们的检测较为复杂,有时难以通过一种通用的检测算法来实现隐写分析,而必须根据不同的嵌入策略进行针对性的分析,并以此为依据设计检测算法。

总体来说,由于视频信息隐藏与视频压缩编码标准的融合导致的不确定性因素太多,因此视频隐写分析比图像隐写分析要复杂许多。若假定已知信息隐藏嵌入点则可以进行针对性的分析,但对于一个陌生的视频文件,隐写分析方通常无法事先获得隐写算法的相关信息。在这种情况下,一方面可以先对待检视频进行隐写嵌入点的判断,进而对该嵌入点进行针对性分析,而特定嵌入点上也可能存在多种信息隐藏技术,这时可以借鉴图像信息隐写分析的方法,同时结合视频隐写独有的特性,设计适当的检测算法;另一方面可以对视频码流中的众多疑似嵌入点进行遍历检测,建立起一个通用的完整的视频隐写分析系统。

9.4.2 数字视频隐写分析典型算法

由于数字视频隐写算法的设计与视频压缩编解码系统相结合,增加了数字视频隐写

分析的难度；而且基于内容的视频信息隐写软件在网络上难于寻找，使得目前视频隐写分析研究尚未充分展开。下面介绍几种典型的数字视频隐写分析方法。

1. 基于质心的数字视频隐写分析方法

J. J. Harmsen 等早在 2003 年就发表了关于视频隐写分析技术的突破性论文。该方案认为隐蔽信息就像是隐藏在载体中的噪声。因此只要是经过隐写操作，载体的性质就会发生变化。而加性噪声的隐藏过程可看作 HCF(Histogram Characteristic Function)被其呈衰减性的质心(Center of Mass，COM)所量化的低通滤波过程。

其中，HCF 是 PMF(Probability of Mass Function)的离散傅里叶变换(Discrete Fourier Transformation，DFT)的结果：

$$H[k] = \mathrm{DFT}[h(n)] = \sum_{n=0}^{N-1} h(n) \cdot \exp\left(-\frac{2\pi jnk}{N}\right) \tag{9-3}$$

作为衡量 HCF 中能量分布的标准，质心的定义如下：

$$C(H[k]) = \frac{\sum_{k \in K} k \cdot |H(k)|}{\sum_{i \in K} |H(i)|} \tag{9-4}$$

对于已知隐写方法的分析过程，只要察看测试视频全局直方统计在傅里叶变换域的质心 HCF-COM 是否小于原始载体即可实现判别。而对于未知隐写机制的分析过程，则定义 Mahalanobis 距离 d^2 并设定阈值进行判别。

$$d^2 = (X - \mathrm{mean})^{\mathrm{T}} \cdot \mathrm{var}^{-1} \cdot (X - \mathrm{mean}) \tag{9-5}$$

理论上 d^2 越大，则表示该视频含密的可能性越高。

2. 基于时域合谋攻击的视频隐写分析方法

Budhia U. 等于 2004 年提出了一种基于时域合谋攻击的数字视频隐写分析方法。在该方法的三个必要的假设下，通过时域一定长度窗口内视频帧的简单线性平均操作得到窗口中心的估计帧 $Z_i(m,n)$，其中 m 和 n 分别代表像素在该帧的 $x-y$ 的位置。该平均操作如下：

$$Z_i(m,n) = \begin{cases} \dfrac{1}{L+1}\sum_{k=1}^{L+1} Xk(m,n), & 1 \leqslant i \leqslant L/2 \\ \dfrac{1}{L+1}\sum_{k=i-L/2}^{i+L/2} Xk(m,n), & L/2 \leqslant i \leqslant N-L/2 \\ \dfrac{1}{L+1}\sum_{k=N-L}^{N} Xk(m,n), & N-L/2 \leqslant i \leqslant N \end{cases} \tag{9-6}$$

分别提取测试帧和估计帧的峰度(Kurtosisi，高斯随机变量的峰度为 0，而该值对其他大多数的分布都是非零的)、熵(Entropy，对于某一指定变量，高斯分布有最大熵)以及其四分位数(25-Percentile，定义为直方图 25%的位置)，并引入相邻帧的相关系数来判定窗口长度的选取是否成功。其中峰度和熵的定义分别如下：

$$\mathrm{Kurtosis} = \frac{E\{[X - E(X)]^4\}}{[E(X - E(X))^2]^2} \tag{9-7}$$

$$\text{Entropy} = -\sum_{\forall i} p_x(i) \cdot \log_2(p_x(i)) \tag{9-8}$$

但是一旦场景发生较大变化，该窗口合谋效果就会失败。同时该攻击是在嵌入率为100%的条件下实现的，实际应用中将不会出现这样的情况。

3. 基于频域的数字视频隐写分析方法

基于 MPEG 系列压缩编码标准的视频资源，是以特定的码流格式进行存储和传输的，码流中包括多种编码元素，例如码头信息、DCT 系数编码信息、运动矢量信息等，信息隐藏算法正是利用这些码流元素中可能存在的随机特性，实现秘密信息的隐蔽嵌入，其中 DCT 系数域及其对应的 VLC 码字域是整个压缩码流中数据量最大的编码元素，因此成为众多视频信息隐藏算法首选的载体数据，最为典型的隐藏算法是中频嵌入算法。

1）针对 DCT 域数字视频信息隐写分析算法模型

苏育挺等提出了基于 DCT 系数能量分布的统计模型，针对不同的视频码流信息进行编码分析，统计各种不同编码图像块中所有 DCT 系数矩阵中各个位置上能量的相对关系，其统计公式如下：

$$\text{SABS}(x,y) = \sum_{n-1}^{N} |\text{DCT}_n(x,y)| \tag{9-9}$$

其中，$\text{DCT}_n(x,y)$表示第 n 块 8×8 DCT 系数矩阵在(x,y)位置处的 DCT 系数值，N 为整个视频码流中的 DCT 系数块数。

通过对 DCT 系数能量分布模型进行分析可以看出，DCT 系数能量数值由低频到高频，沿水平、垂直、对角线方向呈现单调下降趋势，同时，由于整体 DCT 系数分布沿低频到高频呈现广义高斯分布特性，导致其在中频区域呈现凹函数特点，可以利用 DCT 系数能量分布的这一特点构建隐藏分析算法。该算法将为 DCT 系数矩阵中每个系数点设计相应的预测函数 $\text{PSABS}(x,y)$，来检测各系数点呈现的变化是否满足原始码流曲线的单调下降凹函数的特点，具体的预测函数定义如下：

$$\text{PSABS}(x,y) = \min[(P_H(x,y), P_V(x,y), P_D(x,y))] \tag{9-10}$$

式中，$P_H(x,y)$，$P_V(x,y)$，$P_D(x,y)$分别表示水平预测、垂直预测、对角线方向差值预测函数。

$$P_H(x,y) = [\text{SABS}(x-1,y) + \text{SABS}(x+1,y)]/2 \tag{9-11}$$

$$P_V(x,y) = [\text{SABS}(x,y-1) + \text{SABS}(x,y+1)]/2 \tag{9-12}$$

$$P_D(x,y) = [\text{SABS}(x-1,y-1) + \text{SABS}(x+1,y+1)]/2 \tag{9-13}$$

$$\text{diff}(x,y) = [\text{SABS}(x,y) - \text{PSABS}(x,y)]/\text{PSABS}(x,y) \tag{9-14}$$

最后，隐写分析算法根据预测值 $\text{PSABS}(x,y)$与实测数值 $\text{SABS}(x,y)$的相对关系进行判断，如果 $\text{diff}(x,y)$大于阈值 T_p，则判定为疑似嵌入点，否则为正常 DCT 系数点。

该方法依据 DCT 系数块的能量分布统计模型，能够有效地追踪固定 DCT 系数位置出现的能量变化，以此来分析隐写算法引入的可能。但现有的预测器仅是利用能量曲线的单调下降的凹函数特性进行检测，它与该点的真实 DCT 系数能量数值还存在一定的误差。

2）MSU Stego Video 视频隐写分析方法

MSU Stego Video 是由俄罗斯的 MSU Graphics&Media Lab 组织开发并于 2006 年 1 月 31 日公开发布的视频信息隐藏软件，它能够在 AVI 视频文件中隐藏任意类型的信息文件。MSU Stego Video 是真正意义上的视频信息隐藏系统，针对它的视频隐写分析更加具有实用价值。

苏育挺等于 2008 年对 MSU Stego Video 进行了分析。对其输入特殊视频序列，根据结果得出 MSU 是以视频流中的亮度信号 32×32 的块为单元进行嵌入的结论。并且发现每个单元中 4 个 16×16 子块呈棋盘分布，具有交叉调制特征。于是根据它与 MPEG 块编码的差异来提取特征。这属于强针对性隐写分析，它在一定噪声等级和嵌入比率下能够达到一定的识别效果，但通用性不高。

Q. Z. Liu 等针对 MSU 视频隐写软件，提出了一个基于广义 Markov 过程和变换域的联合分布特性的隐写分析方案。建立帧内、帧间及小波近似子带的四个方向联合分布过度矩阵（12 个）。先通过 ANONA（方差分析）得到最优的 496 个特征分量，后送入一个具有 RBF 核的 SVM（支持向量机）进行分类判决。该方法已知隐写机制，不具有通用性，且特征向量庞大，运算复杂。

9.5 数字视频取证技术

9.5.1 数字视频取证技术分类

以数字图像/视频为代表的数字媒体具有易传播、易编辑和易修改等特性，使得普通和专业用户出于各种不同目的，故意修改甚至恶意传播一些经过精心篡改和伪造的数字媒体成为可能。数字媒体取证（digital media forensics）作为信息安全领域的一个新兴研究热点，是指从数字媒体中保持、收集、验证、识别、分析、解释和表示数字证据的科学技术问题。

尽管大多数视频篡改伪造不会引起人们视觉上的怀疑，但是会不可避免地引起视频统计特性的某种变化，从而为数字视频的原始性、真实性和完整性取证提供了可能。此外，通过残留在数字视频内部的捕获设备痕迹以及视频处理的噪声，可以进行数字视频的来源追溯和处理历史恢复。数字视频取证是数字媒体取证的一个重要分支。

1. 视频篡改伪造行为的特点及对视频取证的影响

数字视频具有不同于数字图像的特点，针对数字视频篡改伪造行为所独有的一些特点，应有针对性地发展专用的取证方法。

视频篡改伪造相对于数字图像来说更为复杂和耗时。原因在于：首先，数字视频的数据量更大，且必须尽可能地保持篡改前后的时域一致性（temporal coherency），克服残影（ghost shadow）等问题；其次，视频采集后通常会先进行压缩编码，再以视频流的形式进行存储和传输，而视频编码标准众多，在编码特征工具和码流语法上存在较大的差异；

最后,视频的篡改伪造除了绝大部分为图像篡改伪造手段之外,还有一些视频所特有的篡改伪造手段,这些手段包括重投影(利用摄像设备对已有的视频重新录制)、帧操作(帧插入、帧删除和帧重排序)、超分辨率重建、基于视频对象的视频操纵(包括对象添加、删除和位置变化)等。

相应地,视频被动取证相对于图像取证来说也更为困难,主要体现在:

① 数据量大,对取证算法的计算复杂度的要求高。

② 视频在显示时具有相对较高的帧率,从视觉上难以检测出任何静止的不一致(static inconsistencies)。

③ 数字视频不是简单的比特流,而是具有视觉内容和空间结构内容,并且可能分散在多个设备并以多种格式出现。

④ 对于数字视频所特有的篡改伪造手段,需要进行专门的分析与有针对性的取证。

当然,数字视频取证也存在有利的因素:通过前后连续的多个视频帧,可以得到光照、阴影、深度和遮挡等其他信息,从而可能为视频取证提供更多的线索。

2. 面向真实性鉴别的数字视频被动取证方法

真实性鉴别是指判别数字视频是由成像设备直接捕获的还是遭受了人为的篡改伪造,可能的话还需要确认篡改伪造的区域和程度。根据取证特征的不同,面向真实性鉴别的视频被动取证方法可分为以下两类。

1) 基于视频伪造过程遗留的痕迹的取证

篡改伪造过程会不可避免地会遗留一些痕迹,例如引起视频统计特性的某种变化。该类技术的基本思想是选择那些能够描述伪造痕迹的特征进行取证。这些特征既包括模糊度、块效应和图像区域之间的相似度等图像取证时所用到的特征,也包括 GOP 周期性等视频所特有的特征。目前,该类方法可以对复制-粘贴、MPEG 双重压缩和帧操作等篡改行为进行取证。

(1) 复制-粘贴检测方法的相似度检测

掩盖和去除某些重要目标或者运动对象是一种常见的视频篡改伪造操作。W. H. Wang 提出将视频序列分解为不同的子序列,通过计算每个子序列帧对之间的时间和空间相关矩阵,并与整个视频序列进行比较,判别是否经过帧复制。

(2) MPEG 双重压缩操作的 GOP 特征检测

视频篡改通常会首先对编码后的视频进行还原,篡改操作后再重新编码。因此,篡改后的视频往往经历了双重压缩。双重压缩等视频篡改通常会引入块效应、模糊(blur)和时域抖动(jerkiness)等。这里可以直接将图像双重压缩的篡改检测方法拓展到视频。它结合 MPEG 视频 I、P 和 B 帧的特点,通过计算 MPEG 视频流每个 P 帧的运动误差以及全部帧的平均运动误差,观察运动误差的周期性噪声确定是否发生篡改。

(3) 帧操作检测的时域统计量异常检测

帧操作是常见的视频处理操作,包括帧删除、帧插入和帧重排序。在基于数字水印技术的篡改伪造检测中,帧操作容易检测,因为它会造成水印检测器失去同步,导致水印信息破坏或检测失效。视频被动取证时,则通常利用视频篡改行为会导致前后帧之间时

域统计量的异常,例如帧间预测运动矢量的不一致性等。

(4) 基于数字视频对象操纵的视频修复检测

现有的视频被动取证研究类似于图像被动取证,主要采用某些底层信号特征来进行鉴别。实际上,数字视频不是简单的比特流,而是具有一定空间结构的视觉内容。对于数字视频来说,对象的添加、删除或者修改属于最受关注的恶意篡改与伪造操作,因为视频所包含的视频对象,特别是语义视频对象的改变,往往直接影响人们对视频内容的理解和认识。相对于双重压缩、帧操作等篡改行为,基于视频对象操纵的视频篡改被动取证更有意义。

2) 基于成像设备的一致性进行盲取证

数字视频在捕获成像的过程中,受工作原理和器件物理特性的影响,摄像机的镜头、成像传感器和数字信号后处理都会在成像过程中遗留下特有的设备痕迹和噪声。通过验证设备痕迹和噪声的一致性,可以进行视频真实性的被动取证。目前,这类方法利用的主要特征是传感器的固定模式噪声(Fixed Pattern Noise,FPN)和光子响应非均匀性(Photo-Response Non-Uniformity,PRNU)等。在图像取证中常用的颜色滤波器阵列(Color Filter Array,CFA)的插值方法和相机响应函数(Camera Response Function,CRF)则极少用到视频取证,原因在于 CFA 和 CRF 都是常用的相机特性,而且数字图像和视频的采集成像过程类似。

模式噪声的特点是与所拍摄的场景无关,且在相机的生命期内相对稳定。若将模式噪声视为一个扩频水印,则可以借助水印处理基于相关性检测手段进行判断。其中,FPN 是加性噪声,中高档的拍摄设备可以通过减去一个暗帧进行消除。PRNU 则主要是由于半导体晶片的非均匀性所产生,一般难以消除。因此,通常将 PRNU 模式噪声简称为 PRNU 或者模式噪声。王俊文等提出通过维纳小波滤波器从视频中提取每帧相对稳定的残留模式噪声,并将一段视频中的每帧噪声取平均值从而建立模式噪声。通过比较待鉴别帧的噪声与模式噪声之间的相关性,判别定位篡改区域。

3. 面向来源追溯的数字视频被动取证技术

视频来源取证是指根据视频采集过程、处理过程遗留的痕迹来确定视频捕获设备,甚至设备型号,以追溯视频的来源。对于互联网上篡改伪造视频的非法传播,来源追溯尤为有意义。甚至,数字视频的合法版权所有者也可以借助来源取证技术进行视频拷贝检测。

1) 基于摄像设备内在特性的数字视频来源取证

标准视频文件,例如 AVI 文件等都包含了文件头信息,可以得到捕获设备、采集时间、分辨率和帧率等信息,但是它们容易被修改,不能作为取证的依据。一种可行的方法是提取视频捕获设备内在固有的一些特征。与图像来源取证类似,数字视频的来源辨识依赖于这样的假设:同一设备所获取的视频数据均携带该设备的内在特征,这些特征只与成像管道以及该设备独有的硬件元器件有关,与多媒体数据所表达的内容无关。与真实性取证类似,这类特征包括相机的镜头失真(chromatic aberration)、CCD 的缺陷或者响应不一致引起的传感器模式噪声 PRNU 等。例如利用 PRNU 模式噪声,在离散小波变换的基础上,通过极大似然估计法得到视频序列的 PRNU,并通过正规化相关匹配法

检测 PRNU 的存在，并依据提取的模式噪声进行来源认证。

2）利用数字视频码流特征进行来源取证

数字视频来源取证还可以借助输出数据流的统计特征进行。视频编码标准通常只规定了编码的框架、特征工具和解码器比特流的句法结构等，而编码器的实现具有相当大的灵活性。因此，不同的商家采用了不同的速率控制方案，每帧输出的码流会在码率的分布控制上有明显的差异。甚至，不同的运动估计算法，编码器采用不同的匹配准则、搜索路径等，都可能为视频来源取证提供依据。例如以一个 GOP 作为训练样本，提取了两类码流的特征并用支持向量机进行训练分类，然后判决待测视频序列来源哪种类型的 MPEG 编码系统设备。

9.5.2 数字视频取证技术典型算法

基于数字视频的取证技术在研究深度、广度及技术成熟度等方面都远滞后于数字图像取证技术，公开发表的学术论文也较少。但其日益迫切的需求和巨大的市场潜力，促使许多具有静止图像取证研究背景的科研团队将目光转向了数字视频。下面简单介绍一些数字视频取证的典型算法。

1. 帧复制或插入、删除操作的检测

在数字视频的篡改操作中，帧复制或插入、删除操作是最简单也是最常用的视频伪造手段。同一视频序列帧复制篡改方法也有较大的破绽，其最直接的检测方法就是全局搜索，但这种穷举搜索取证技术的缺点是计算量大，对自然噪声的鲁棒性差。H. Farid 提出了利用视频序列的时间和空间相关性来找到复制帧，具体来说，首先将一个视频序列划分为许多的互相重叠的子序列，然后计算每两两子序列的时间相关性，大于判决门限的帧则标记为可疑帧，最后计算可疑帧的每个子块的空间相关性，如果结果大于判决门限，则判该帧为复制帧。该方法的检测性能高，但执行速度较慢。

局部帧复制的篡改检测方法也可以采用 H. Farid 提出的块相关性检测方法，但其匹配需要时间，执行速度相对较慢；而 Chih-Chung Hsu 提出利用噪声残差的相关性来定位篡改区域的方法，不仅速度快，而且还能抵御常见的局部修复方法——基于样本的纹理合成。Hsu 首先利用小波降噪滤波器过滤原始序列获得了噪声残差，然后把每帧划分为互不重叠的 $N \times N$ 块，并计算每相邻两帧对应空间位置块之间的相关性，接着利用简单门限法做一次粗分类，最后用最大期望的方法（Expectation-Maximization，EM）估计出高斯混合模型（Gaussian mixture model，GMM）的参数，根据估计的参数，使用贝叶斯分类器找到最优门限值。该方法不需要预先得到噪声残差的统计特性，检测效果好，但对太亮或太暗的区域会产生虚警，因为这些块的噪声残差能量都比较小，容易受量化噪声的影响。

针对帧插入和帧删除的篡改操作，Min Wu 提出了利用视频序列块效应时域模型的检测方法。MPEG 压缩会在不同类型的视频帧中引入不同的块效应，在给定了 GOP 结构的情况下，块效应的强度是一个随时间变化的常规模型函数。如果插入或删除了一个 MPEG 视频文件的某几帧，接着重新压缩成 MPEG 文件，则第一次压缩引入的块效应仍

然存在，且会依据删帧或插帧数目和第一次压缩 GOP 结构类型的不同，对第二次压缩引入的块效应平均强度造成不同影响。通过提取 MPEG 视频的特征曲线的不一致性，检测出 GOP 结构的变化，从而揭示出篡改操作。该方法对视频内容的变化不敏感，鲁棒性高，且容易结合其他特征量如预测误差，从而提出更全面的取证算法。

2. 模糊检测

在伪造篡改者经过复制-粘贴篡改操作后，为了消除伪造视频上的篡改痕迹，往往会对篡改部分进行如缩放、旋转、模糊预先处理和润饰操作。在对篡改部分进行缩放、旋转操作后，会在篡改视频中留有重采样的痕迹，尽管这种痕迹不会引入视觉上的差别，但由于插值（上采样或下采样）的原因，经重采样后篡改视频块的像素与其周围像素之间会产生周期性的关系。Zhang 提出利用运动物体的轨迹和运动前景的不一致性来检测出 ghost 模糊效应。首先通过块匹配方法将视频每帧分为运动前景和背景两部分，并建立运动前景马赛克模型。利用连续帧差信号和数学形态模型计算出运动物体的轨迹。如果运动前景的马赛克模型与运动轨迹相一致，则判该序列为完整的，否则，判该视频序列经过了模糊篡改操作。

3. MPEG 双重压缩检测

篡改后的视频一般都是像素域的，所以需要重新压缩一次。如果原始视频是 MPEG 格式，修改后仍旧保存为 MPEG 格式，则视频序列就经过了 MPEG 二次压缩操作，这是一种不可逆的有损压缩过程。由于 MPEG 二次压缩操作对视频数据进行了两次量化，引入了一次 MPEG 压缩所没有的特征，通过统计检测很容易发现视频是否经过了 MPEG 二次压缩。当然，MPEG 二次压缩并不说明视频一定就经过了篡改，有些未篡改的视频为了节省存储空间也会引入 MPEG 二次压缩。

视频可以看成由一系列的图像组成，因而视频二次压缩检测方法也就可以借鉴图像领域的检测方法。W. H. Wang 提出了 MPEG 二次压缩检测方法，指出视频二次压缩会引入空域痕迹和时域痕迹。在空域中，一个视频序列被分为一系列的图像，并把 MPEG 二次压缩码流中的 I 帧看做是一个经过 JPEG 二次压缩的图像，因此，JPEG 二次压缩检测方法可以直接扩展到视频编码系统中；在时域中，当有帧插入或删除时，连续几个帧的运动估计误差的分布会呈现周期性，这个特点也可以用来鉴别 MPEG 二次压缩。但是该方法有很多的限制条件，比如需要一帧中所有宏块的量化参数都是一样的，且第二次压缩的量化参数要比第一次压缩的小等。在实际的编码系统中，一般只能改变其输出码率，而不允许直接修改量化参数。因此，该方法离实际的应用还有很大的距离。

Yun Q. Shi 发现经过一次压缩编码后的视频宏块的第一个非零的量化 AC 系数的分布服从广义的第一定律，其公式如下：

$$P(x) = N\log10\left(1+\frac{1}{s+x^q}\right),\quad x=1,2,\cdots,9 \tag{9-15}$$

x 为第一个非零的量化 AC 系数的值，N、s、q 为精确描述其分布曲线的参数。而 MPEG 二次压缩操作将打破该分布，使得第一个非零的量化 AC 系数的分布不再服从第

一定律。通过计算曲线的三个拟合性能评价参数：平方和误差（Sum of Squares Due to Error，SSE），相关系数（Coefficient of Multiple Determination，R-square），均方根误差（Root Mean Squared Error，RMSE），以及归一化概率分布曲线的9个值，组成了12维的特征量，每种类型帧都能提取12维特征，所以一共36维特征。以图像组（GOP）为检测单元，针对多个P、B帧的情况，先计算每帧12维特征，然后平均每种类型帧的特征向量，最后再联合I帧的12维特征一起输入到SVM中进行判决。如果判为经过了二次压缩的GOP个数占整体的百分比超过了自定义门限时，就判该视频序列经过了二次压缩操作，否则判为原始压缩序列。该方法能够适应变码率（VBR）和定码率（CBR）两种情况，且性能较佳。但该方法比较适应于前后两次为同一编码器的情况，且对帧错位操作非常敏感。

4. 数字视频来源取证方法

目前有很多种数字设备都可以生成数字视频，视频来源认证就是要在不明视频来源的情况下，单从数字视频本身来判别它的生成设备。不同摄取设备来源的数字视频虽然在视觉上区别不大，但由于各种视频生成设备特征的不同（如镜头、感光器件），其生成的数字视频也会有不同的特征，现有的视频来源认证就是通过提取这些能够区别视频来源的特征，建立特征库，对数字视频的来源进行盲认证。

已有的取证算法大多利用传感器的缺陷来鉴别视频其来源。比较经典的方法是Kurosawa提出的CCD芯片的暗电流不一致性。这个方法假设CCD中某些像素的暗电流产生率会偏离平均值，而这些缺陷像素造成了一个固定的模式噪声，它对一个单独的摄像机来说是独一无二的，利用它便可鉴别视频其来源。而M. Chen扩展他们的图像定向技术到视频领域，并提出利用数字传感器的响应不一致性PRNU来鉴别数字摄像机来源。由于硅片的不均匀性以及摄像机制造过程的不完美性，造成像素传感器对光具有不同的敏感度。这种属性不随时间变化，与图像传感器是一一对应的。M. Chen的方法是利用最大似然估计从帧序列中估计出PRNU，并利用归一化互相关函数来检测PRNU的存在。Wiger van Houten也是利用摄像机的PRNU函数来鉴别视频来源，但不同的是他提出一种基于小波技术的PRNU提取方法。

思　考　题

9.1　什么是4∶2∶2标准和4∶2∶0标准？在制定这两个标准时考虑到了哪些因素（依据）？

9.2　对于数字RGB坐标中的下列彩色，确定它们在YCbCr坐标中的值。

(1) (255,255,255)；

(2) (0,255,0)；

(3) (255,255,0)；

(4) (0,255,255)。

9.3　假设一个8位灰度级图像，如何估算出该图像的熵？

9.4　数字视频水印与传统图像水印技术有哪些不同点？

9.5 试比较完全加密和选择性加密的优缺点。

9.6 简述基于 LSB 数字视频技术的信息隐藏与提取。

9.7 举例说明数字媒体的知识产权保护问题(至少三个例子)。

参考文献

[1] 毕厚杰,五健. 新一代视频压缩编码标准——H.264/AVC. 北京:人民邮电出版社,2009.

[2] 陈威兵,杨高波,陈日超,等.数字视频真实性和来源的被动取证. 通信学报,2011,32(6):177-183.

[3] 裴智勇,张春红. H.26x 与 MPEG-x. 电信工程技术与标准化,2005.

[4] 施昌林. 视频隐写分析算法研究. 上海交通大学硕士论文,2009.

[5] 岳斌. 基于 H.264 的视频内容安全技术研究. 北京机械工业学院硕士论文,2008.

[6] 廉士国,孙金生,王执铨. 视频加密算法及其发展现状. 信息与控制,2004,33(5):560-566.

[7] Wee S J, Apostolopoulos J G. Secure scalable video streaming for wireless networks. In: Proceedings of the IEEE International Conference on Acoustics,Speech,and Signal Processing,Salt Lake City,2001. 2049-2052.

[8] Zeng W J, Lei S M. Efficient frequency domain selective scrambling of digital video. IEEE Transactions on Multimedia,2003,5(1):118-129.

[9] Noorkami M,Mersereau R M. A framework for robust watermarking of H.264-encoded video with controllable detection performance. IEEE Transactions on Information Forensics and Security, 2007,2(1): 14-23.

[10] Tian L,Zheng N,Xue J,et al. A CAVLC-based blind watermarking method for H.264/AVC compressed video. In: Proceedings of IEEE Asia-Pacific Services Computing Conf., 2008, 1295-1299.

[11] Noorkami M, Mersereau R M. Compressed-domain video watermarking for H.264. In: Proceedings of IEEE International Conference on Image Processing,2005,890-893.

[12] Richardson I E. H.264 and MPEG-4 Video Compression: Video Coding for Next-Generation Multimedia. Hoboken,NJ: Wiley,2003.

[13] Petitcolas F A P,Anderson R J,Kuhn M G. Information Hiding-A Survey. Proceedings of IEEE, 1999,87(7):1062-1078.

[14] Koch E,Zhao J. Towards robust and hidden image copyright labeling. In: Proceedings of IEEE Workshop on Nonlinear Signal and Image Processing,Neos Marmaras,Greece,1995,452-455.

[15] A Popescu, Farid H. Exposing digital forgeries by detecting traces of re-sampling. IEEE Transactions on Signal Processing. 2005,53(2): 758-767.

[16] Su Y T,Zhang J,Ji Z. A Source Video Identification Algorithm Based on Features in Video Stream. International Workshop on Education Technology and Training& International Workshop and International Workshop on Geosciences and Remote Sensing,2008,719-723.

[17] Zhang J,Maitre H. Embedding watermarking in MPEG video sequence. IEEE Fourth workshop on Multimedia signal Processing,2001,535-540.

[18] Hartung F,Girod B. Digital Watermarking of raw and compressed video. In: Proceedings. Of SPIE. 1996,2952:205-213.

[19] 王俊文,刘光杰,等.基于模式噪声的数字视频篡改取证.东南大学学报(自然科学版),2009,

38(A02):13-17.

[20] 秦运龙,孙广玲,张新鹏. 利用运动矢量进行视频篡改检测. 计算机研究与发展,2009,46(SUPPL):227-233.

[21] 熊潇,黄征,徐彻,等. 基于预测残差检测的数字视频篡改鉴定. 信息安全与通信保密,2008,5(12):128-130.

[22] 周琳娜,王东明. 数字图像取证技术. 北京:北京邮电大学出版社,2008.

[23] Spanos G A, Maple T B. Security for real-time MPEG compressed video in distributed multimedia applicatior. Computers and commurnication,1996.

[24] Changgui Shi, Bharat Bhargava. A fast MPEG Video encryption algorithm. In: Proceedings of the 6th ACM international Multimedia Conference. 1998: 81-88.

[25] Tang L. Methods for encryptior and decryptior MPEG Video data efficiently. In: Proceedings of the 4th ACM International Multimedia Conference. 1996: 219-230.

[26] 苏育挺,张承乾. 一种 DCT 域视频信息隐藏分析算法. 哈尔滨工业大学学报. 2006.

[27] 徐俊瑜. 数字视频被动取证技术研究. 天津大学硕士学位论文,2010.

[28] 张承乾. 视频信息隐藏分析研究. 天津大学博士论文,2008.

[29] Simnrons G J. The prisoners' Problem and the Subliminal Channel. In: Proceedings of CRYPTO'1983, 1984: 51-67.

第 10 章 chapter 10

数据库安全

本章学习目标

随着信息化建设的深入，数据库在各种信息系统中得到了广泛的应用，其安全问题也日益突出。本章将介绍数据库的基本特性、数据库所面临的安全威胁，以及当前的数据库安全技术。具体包括数据库的机密性、完整性、访问控制以及安全管理等方面的知识。

通过本章的学习，应掌握以下内容：

(1) 数据库安全的基本概念以及数据库面临的安全威胁。

(2) 数据库访问控制技术。

(3) 数据库水印技术。

(4) 数据库安全管理：数据库加密、审计等。

数据库和数据库技术在不断增长的计算机应用中起着越来越大的作用。在计算机应用的各个领域，数据库都起着至关重要的作用。如果数据库的安全没能得到有效保护，计算机和网络应用的深度和广度都将大受影响。对于数据库的用户来说，机密性和完整性都非常重要。

当前，数据库安全问题已开始引起人们的关注。虽然已经开发了一些数据库安全技术，但仍然还存在一些无法控制的安全隐患。本章将从数据库的基本概念入手，介绍数据库的相关特性，分析当前数据库所面临的安全问题，对常见的数据库安全技术，如数据库安全访问控制技术、数据库数字水印技术、数据库加密技术以及数据库安全管理技术进行介绍。

10.1 数据库安全基本概念

10.1.1 数据库的基本概念

数据库(DataBase, DB)是在数据库管理系统(DataBase Management System, DBMS)的集中控制下，按一定的组织方式存储起来的、相互关联的数据集合，能为多个用户共享，且具有数据冗余度小、独立性和安全性高等特点。数据库中的数据独立于使用

数据的程序,对数据的增加、删除、修改和检索等操作都由 DBMS 进行统一管理和实现。目前最常用的数据库是关系型数据库。随着数据库技术的发展,涌现出了许多新型数据库,例如分布式数据库、多媒体数据库和数据仓库等。

数据库系统主要包括两个核心:一个是按一定规则组织的数据集合本身;另一个是 DBMS,它为用户提供访问接口并且具有数据库的管理、维护功能,保证数据库的安全性、可靠性和完整性。数据库支撑示意图如图 10-1 所示。

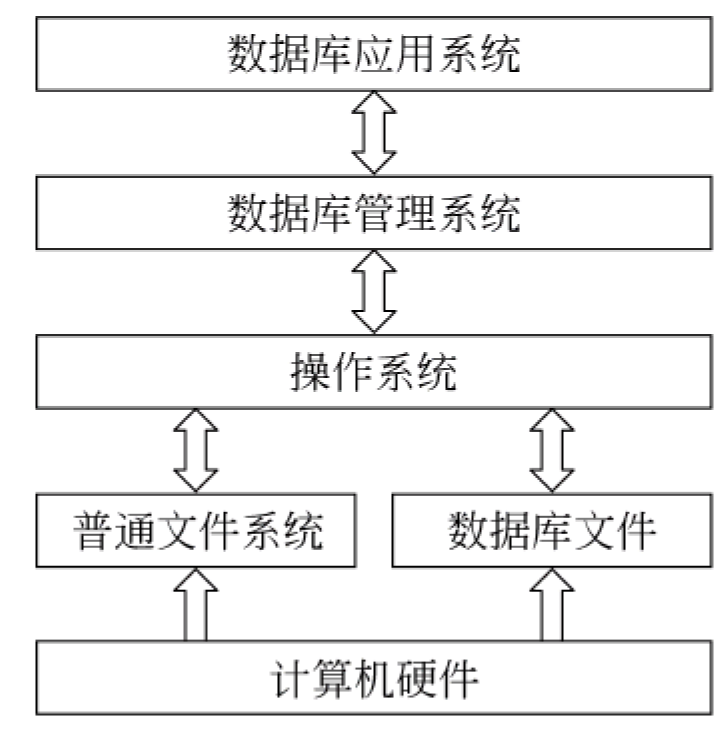


图 10-1 数据库支撑示意图

数据库除了具有多用户、高可靠性、频繁的更新和数据文件大等特性外,还具有数据共享、减少数据冗余、数据的一致性、数据的独立性、数据的保密性、数据的完整性、并发控制和故障恢复等技术特性。

数据库文件由记录(record)组成,每个记录包含了一组相关的数据。如表 10-1 所示,在一个名字地址文件中,每个记录由名字和地址数据组成。每个记录都包含域(field)或元素(element),即它们的基本数据项。

表 10-1 一个数据库的实例

ADAMS	212 Market St.	Columbus	OH	43210
BENCHLY	501 Union St.	Chicago	IL	60603
CARTER	411 Elm St.	Columbus	OH	43210

数据库的逻辑结构称为模式(schema)。一名特殊的用户可能只允许访问数据库的一部分,称之为子模式(subschema)。通过模式或子模式,数据库可以只显示用户想看到或需要看到的元素。

数据库的规则要求以列名识别列,列名也称为数据库的属性(attribute),列的集合构成了一个关系(relation)。关系描述了有关数据值的簇(cluster),大多数来源于对人类关系的描述。

用户通过使用 DBMS 的命令与数据库管理器交互,这些命令有:检索、修改、增加或删除数据库中的域和记录。

10.1.2 常用数据库系统与 SQL 语言

1. 常用的数据库系统

常用的数据库系统包括 DB2、Oracle、Informix、Sybase、SQL Server、PostgreSQL 和 mySQL 等。

1) DB2

IBM 于 1997 年完成了 System R 系统的原型,1980 年开始提供集成的数据库服务器——System/38,随后是 SQL/DS for VSE 和 VM,其初始版本与 System R 研究原型

密切相关,DB2 for MVSV1 在 1983 年推出。该版本的目标是提供这一新方案所承诺的简单性,数据不相关性和用户生产率。1988 年推出的 DB2 for MVS 提供了强大的在线事务处理(OLTP)支持,1989 年和 1993 年分别以远程工作单元和分布式工作单元实现了分布式数据库支持。最近推出的 DB2 Universal Database 6.1 则是通用数据库的典范,是第一个具备网上功能的多媒体关系数据库管理系统,支持包括 Linux 在内的一系列平台。

2) Oracle

Oracle 公司前身为 SDL 公司,由 Larry Ellison 和另两个编程人员在 1977 创办。1979 年,Oracle 公司引入了第一个商用 SQL 关系数据库管理系统。Oracle 公司是最早开发关系数据库的厂商之一,其产品支持最广泛的操作系统平台。目前 Oracle 关系数据库产品的市场占有率名列前茅。

3) Informix

Informix 公司于 1980 年成立,目的是为 UNIX 等开放操作系统提供专业的关系型数据库产品。公司的名称 Informix 便是取自 Information 和 UNIX 的结合。Informix 第一个真正支持 SQL 语言的关系数据库产品是 Informix SE(StandardEngine)。InformixSE 是在当时的微机 UNIX 环境下主要的数据库产品。它也是第一个被移植到 Linux 上的商业数据库产品。

4) Sybase

Sybase 公司成立于 1984 年,公司名称 Sybase 取自 system 和 database 相结合的含义。Sybase 公司的创始人之一 Bob Epstein 是 Ingres 大学版(与 System/R 同时期的关系数据库模型产品)的主要设计人员。公司的第一个关系数据库产品是 1987 年 5 月推出的 Sybase SQL Server 1.0。Sybase 首先提出 Client/Server 数据库体系结构的思想,并率先在 Sybase SQL Server 中实现。

5) SQL Server

1987 年,微软公司和 IBM 公司合作开发完成 OS/2,IBM 在其销售的 OS/2 ExtendedEdition 系统中绑定了 OS/2Database Manager,而微软产品线中尚缺少数据库产品。为此,微软将目光投向 Sybase,同 Sybase 签订了合作协议,使用 Sybase 的技术开发基于 OS/2 平台的关系型数据库。1989 年,微软发布了 SQL Server 1.0 版。

6) PostgreSQL

PostgreSQL 是一种特性非常齐全的自由软件的对象——关系性数据库管理系统(ORDBMS),它的很多特性是当今许多商业数据库的前身。PostgreSQL 最早开始于 BSD 的 Ingres 项目。PostgreSQL 的特性覆盖了 SQL-2/SQL-92 和 SQL-3。首先,它包括了可以说是目前世界上最丰富的数据类型的支持;其次,目前 PostgreSQL 是唯一支持事务、子查询、多版本并行控制系统、数据完整性检查等特性的唯一的一种自由软件的数据库管理系统。

7) MySQL

MySQL 是一个小型关系型数据库管理系统,开发者为瑞典 MySQL AB 公司。在 2008 年 1 月被 Sun 公司收购。目前 MySQL 被广泛地应用在 Internet 上的中小型网站

中。由于其体积小、速度快、总体拥有成本低，尤其是开放源码这一特点，许多中小型网站为了降低网站总体拥有成本而选择了 MySQL 作为网站数据库。

2. SQL 语言

SQL 是 Structured Query Language（结构化查询语言）的简写。是一种高级的非过程化编程语言，允许用户在高层数据结构上工作。它不要求用户指定对数据的存放方法，也不需要用户了解具体的数据存放方式，所以具有完全不同底层结构的不同数据库系统，可以使用相同的 SQL 语言作为数据输入与管理的接口。它以记录集合作为操作对象，所有 SQL 语句接受集合作为输入，返回集合作为输出，这种集合特性允许一条 SQL 语句的输出作为另一条 SQL 语句的输入，所以 SQL 语句可以嵌套，这使其具有极大的灵活性和强大的功能，在多数情况下，在其他语言中需要一大段程序实现的功能只需要一个 SQL 语句就可以达到目的，这也意味着用 SQL 语言可以写出非常复杂的语句。

SQL 最早是 IBM 的圣约瑟研究实验室为其关系数据库管理系统 SYSTEM R 开发的一种查询语言，它的前身是 SQUARE 语言。SQL 语言结构简洁，功能强大，简单易学，所以自从 IBM 公司 1981 年推出以来，SQL 语言得到了广泛的应用。如今无论是像 Oracle、Sybase、Informix、SQL Server 这些大型的数据库管理系统，还是像 Visual Foxpro、PowerBuilder 这些 PC 上常用的数据库开发系统，都支持 SQL 语言作为查询语言。

1992 年，ISO 和 IEC 发布了 SQL 国际标准，称为 SQL-92。ANSI 随之发布的相应标准是 ANSI SQL-92。ANSI SQL-92 有时被称为 ANSI SQL。尽管不同的关系数据库使用的 SQL 版本有一些差异，但大多数都遵循 ANSI SQL 标准。SQL Server 使用 ANSI SQL-92 的扩展集，称为 T-SQL，其遵循 ANSI 制定的 SQL-92 标准。

SQL 语言包含 4 个部分：

（1）数据定义语言（DDL）用于定义和管理对象。例如，CREATE、DROP、ALTER 等语句。

（2）数据操作语言（DML）用于操作数据库对象所包含的数据。例如，INSERT（插入）、UPDATE（修改）、DELETE（删除）等语句。

（3）数据查询语言（DQL）用于数据库的查询。例如，SELECT 语句。

（4）数据控制语言（DCL）用于控制用户对数据库对象操作的权限。例如，GRANT、REVOKE、COMMIT、ROLLBACK 等语句。

10.1.3 数据库的数据特点

与多媒体数据相比，关系数据库中的关系数据的主要区别在于：

（1）多媒体数据对象是由大量的位组成的，并且许多位是冗余的。关系数据库则是由许多独立的元组组成，每个元组代表一个单独的对象，数据间一般存在依赖关系，难以找到可辨认的冗余空间。

（2）多媒体数据对象各个点之间主要存在空间上的有序关系。而组成关系数据库的元组之间以及元组的属性值集合之间是无序的。

(3) 多媒体数据对象某个部分的删除或替换,很容易引起知觉上的变化,而关系数据库却可以简单地去掉一些元组或者用其他类似的关系数据中的元组来代替而不易被发觉。这使得数据库水印易于被攻击且难以发现。

(4) 数据库数据主要被机器程序读取和处理,无法像多媒体数据那样基于人类视觉模型(HVS)或听觉模型(HAS)来实现数字水印的隐蔽嵌入。

(5) 静态的多媒体数据很少进行更新,而数据库一般更新频繁。

10.1.4 数据库安全概述

数据库安全是指数据库的任何部分都不允许受到恶意侵害或未经授权的存取或修改,以保证数据的安全可靠和正确有效。

随着计算机资源共享和网络技术应用的日益广泛,越来越多的数据库需要通过网络进行存储和发布。对于一些重要的部门、有价值的数据,如网上银行、购物、证券等部门的用户资料数据库,这些数据往往蕴涵巨大的社会价值与经济价值,因而会经常成为不法分子感兴趣的目标,极易遭到攻击和破坏,因此需要保护数据的安全性和完整性。此外,一些数据库应用需要将数据库产品出售给客户(如地理信息系统中一般就包含价格不菲的空间数据库),一些数据库业务(如数据挖掘等)需要向合作伙伴提供完整的数据,这些都需要严格的数据库版权保护措施。数据库版权保护是数据库安全的另一项重要作用。

因此,数据库安全主要包括三个方面的内容:保密性、完整性和可用性。

- 保密性:不允许未经授权的用户存取信息。
- 完整性:只允许被授权的用户修改数据。
- 可用性:不应拒绝已授权的用户对数据进行存取。

关于数据库安全,以 C. P. Pfleeger 在 Security in Computing Database Security 论文中的定义最具有代表性。该定义从以下方面对数据库安全进行了描述:

(1) 物理数据库的完整性。数据库中的数据不被各种自然的或物理的问题所破坏。如水灾、火灾、电力问题造成的硬件故障或设备故障等,会导致数据库的损坏和丢失。

(2) 可信计算基。可信计算基(Trusted Comptuing Base,TCB)是实现数据库安全的所有实施策略与机制的集合。它是实施、检查、监督数据库安全的一种抽象机构。

(3) 逻辑数据库的完整性。对数据库结构的保护,如对其中一个字段的修改不应该破坏其他字段。逻辑上的威胁主要是指对信息未被授权的存取,可以分为三类:信息泄露,包括直接和非直接(通过推理)地对保护数据的存取;非法的数据修改,由操作人员的失误或非法用户的故意修改引起;拒绝服务,通过独占系统资源导致其他用户不能访问数据库。为了消除逻辑上的威胁,DBMS 必须提供可靠的安全策略,以确保数据库的安全性。

(4) 元素安全性。存储在数据库中的每个元素都是正确的。当数据库被使用时,应确保合法用户得到正确的数据。数据库不仅储存数据,还要为使用者提供信息。应该确保合法用户在一定规则的控制和约束下使用数据库,同时应当防止入侵者或非授权者非法访问数据库。

(5) 可审计性。可以追踪存取和修改数据库元素的用户,同时能对各种安全性事件进行检查、跟踪和记录。它提供了信息系统安全事件的证明和依据。

(6) 访问控制。确保只有授权的用户才能访问数据库,这样不同的用户被限制在不同的访问方式。

(7) 身份验证。不管是审计追踪或者是对某一数据库的访问都必须要经过严格的身份验证。它是一种鉴别某一实体身份真伪性的技术,是防止冒充攻击的重要手段。

(8) 可用性。可用性是指对授权的用户可以访问数据库中的授权数据和一般数据的能力。DBMS既是程序也是系统,用户通常把DBMS看做是用来执行特殊任务的基本工具。但是,当系统不可用时(忙于为其他用户服务、正被维护或更新时),用户会清楚地意识到此时DBMS是不可用的。例如,两个用户同时要求访问某一个记录时,DBMS必须做出决策:哪一个用户暂时不能访问该记录。有时,DBMS必须限制访问某些不受保护的数据以免泄露需要保护的数据,而不管用户是否愿意。

(9) 数据库中的数据加密。加密是信息安全中的一种传统方法,在数据库中也不例外,但是由于数据库中数据结构的特殊性以及数据操纵的要求,使得对数据库中数据加密有别于其他信息安全领域中的加密。此外,对数据库中数据进行加密后会对数据操纵与控制造成一定的影响。因此,一般对数据库的加密要慎重行事。

(10) 数据库安全的三权分立模式。在数据库管理系统中,DBA(Database Administrator)具有至高的权力,但是为保证数据的安全,需要将DBA的权力作重新调整,这就是所谓的三权分立模式,在该模式中DBMS由三部分高级别人员管理:

① DBA:具有管理DBMS的最高权力,但有关数据安全的管理权力除外。

② SA(Secure Administrator):即"安全管理员",它具有管理数据库中数据安全的最高权力,但有关审计管理权力除外。

③ AT(Auditor):即"审计员",由于审计在数据库安全中的特殊作用,须设置专门的审计员以负责数据库安全中的审计管理工作。

数据库安全负责信息存取安全中数据库这个层次的安全,具体来说,就是保证用户正确地访问数据库,防止非法访问数据库。为实现数据库的存取安全,需要解决的技术问题主要包括:

① 存取控制模型:为保证数据库存取的安全需对数据库访问建立一定的控制机制,称为"数据库的存取控制",而这种存取控制的抽象结构称为"存取控制模型"。目前有多种存取控制模型能适应多种不同的应用,但每种模型也均存在一些不足,因此需要研究新的存取控制模型适应不同需求。

② 语义推理技术:由于数据库中数据之间往往存在着语义上的关联,因此在某些情况下,一些用户可以从它有权访问的数据中通过语义关联推导出它无权访问的数据来,对该问题的研究可以找出数据库中的推理通道,它们是以隐蔽形式出现的也可称为隐蔽通道。找出推理通道的目的是为了堵塞这些通道以防止非法访问数据。

③ 数据库中数据加密技术:数据加密一直是信息安全的一项重要技术,对数据库安全也是如此,在数据库领域中由于其数据的特殊性,使得一般数据的加密存在不少困难,因此需要研究数据库中数据加密的原理与方法。

由上述内容可知，数据库安全的目标是保证对数据的正确访问与防止对数据的非法访问，数据库安全包含的内容为对数据库中存取控制模型、语义推理技术、数据加密技术以及数据库标准规范的制定。

10.1.5 数据库安全标准

目前，国际上及我国均颁布了有关数据库安全的等级标准，最早的标准是美国国防部(DOD)在 1985 年颁布的“可信计算机系统评估标准”(Trusted Computer System Evaluation Criteria，TCSEC)。1996 年国际标准化组织 ISO 颁布了“信息安全技术——信息技术安全性评估准则”(Information Technology Security Techniques——Evaluation Criteria for IT security)，简称 CC 标准。我国政府于 1999 年颁布了“计算机信息系统评估准则”。目前国际上广泛采用的是美国标准 TCSEC，在此标准中将数据安全划分为 4 组 7 级，我国标准则划分为五个级别。

1. TCSEC(TDI)标准

TCSEC(TDI)标准是目前常用的标准，在此标准中将数据库安全分为 4 类 7 级。

(1) D 级标准。为无安全保护的系统。

(2) C1 级标准。满足该级别的系统必须具有如下功能：

① 主体、客体及主、客分离。

② 身份标识与鉴别。

③ 数据完整性。

④ 自主访问控制。

其核心是自主访问控制。C1 级安全适合于单机工作方式，目前国内使用的系统大都符合此标准。

(3) C2 级标准。满足该级别的系统必须具有如下功能：

① 满足 C1 级标准的全部功能。

② 审计。

(4) B1 级标准。满足该级别的系统必须具有如下功能：

① 满足 C2 级标准全部功能。

② 强制访问控制。

(5) B2 级标准。满足该级别标准的系统必须具有如下功能：

① 满足 B1 级标准全部功能。

② 隐蔽通道。

③ 数据库安全的形式化。

一个数据库系统凡是符合 B1 级标准的都称为安全数据库系统(secure DB system)或可信数据库系统(trusted DB system)。目前我国国内所使用的系统基本不是安全数据库系统。

(6) B3 级标准。满足该级别的系统必须具有如下功能：

① 满足 B2 级标准的全部功能；

② 访问监控器。

(7) A 级标准。满足该级别的系统必须具有如下功能：

① 满足 B3 级标准的全部功能；

② 较高的形式化要求。

此级为安全之最高等级，应具有完善的形式化要求，目前尚无法实现。

2. 我国标准与 TCSEC(TDI)标准对比

我国国家标准于 1999 年颁布，为与国际接轨其基本结构与 TCSEC(TDI)标准相似。我国标准 5 级，从第 1 级到第 5 级基本上与 TCSEC(TDI)标准的 C(C1,C2)级及 B 级(B1,B2,B3)一致，我国标准与 TCSEC(TDI)标准比较如表 10-2 所示。

表 10-2 TCSEC 标准与我国标准的比较

TCSEC 标准	我 国 标 准	TCSEC 标准	我 国 标 准
D 级标准	无	B2 级标准	第 4 级：结构化保护级
C1 级标准	第 1 级：用户自主保护级	B3 级标准	第 5 级：访问验证保护级
C2 级标准	第 2 级：系统审计保护级	A 级标准	无
B1 级标准	第 3 级：安全标记保护级		

10.2 数据库面临的安全威胁

随着数据库应用越来越广泛，其安全隐患也越来越多，当前数据库主要存在十类安全威胁。

1. 权限滥用

用户(或应用程序)应只能在自己的工作职责范围内对相应的数据进行访问，一旦被授予超出了其工作职能所需的数据库访问权限时，这些权限就可能被恶意滥用。例如，一个银行职员在工作中只需要能够更改客户的联系信息，不过他可能会利用过高的数据库更新权限来更改客户的存款金额。

2. 合法权的滥用

合法的数据库权限被用于未经授权的目的。假设一个医务人员拥有可以通过 Web 应用程序查看某个患者病历的权限。通常情况下，该 Web 应用程序的结构限制用户只能查看单个患者的病史，但是，恶意的医务人员可以通过使用其他客户端(如 Excel)连接到数据库来规避这些限制。通过使用 Excel 以及合法的登录凭据，该医务人员就可以检索和保存所有患者的病历。这种私自复制患者病历数据库的副本的做法是不符合任何医疗组织的患者数据保护策略的。

3. 权限设定

任何系统都有可能存在漏洞，数据库管理系统也是如此。攻击者可以利用数据库平

台软件的漏洞将普通用户的权限转换为管理员权限，达到不可告人的目的。漏洞可以在存储过程、内置函数、协议实现甚至是 SQL 语句中找到。例如，一个金融机构的软件开发人员利用有漏洞的函数来获得数据库管理权限，恶意的开发人员可以使用管理权限禁用审计机制、开设伪造的账户以及转账等。

4. 平台漏洞

底层操作系统（Windows 2000、UNIX 等）中的漏洞和安装在数据库服务器上的其他服务中的漏洞可能会导致未经授权的访问、数据破坏或拒绝服务。例如，“冲击波病毒”就是利用了 Windows 2000 的远程过程调用协议漏洞为拒绝服务攻击创造条件。

5. SQL 注入

SQL 注入攻击主要是由于程序设计中忽略了 SQL 语句检查引起的。在 SQL 注入攻击中，入侵者通常将未经授权的数据库语句插入（或“注入”）到有漏洞的 SQL 数据信道中。通常情况下，攻击所针对的数据信道包括存储过程和 Web 应用程序输入参数，这些注入的语句被传递到数据库中并在数据库中执行。使用 SQL 注入，攻击者可以不受限制地访问整个数据库。在实际应用中，可以将以下三个技术结合使用来抵御 SQL 注入：入侵防御系统（IPS）、查询级别访问控制和事件相关。

6. 审计记录缺陷

自动记录所有敏感的和/或异常的数据库事务应该是所有数据库部署基础的一部分。如果数据库审计策略不足，则数据库将在很多级别上面临严重风险。CSDN、天涯等社区的帐号被泄露后，数据库审计被广泛关注。

7. 拒绝服务

拒绝服务（DoS）是一个宽泛的攻击类别，其基本原理是攻击时利用合理的服务请求占用过多的资源，导致正常用户的访问被拒绝。可以通过多种技巧为 DoS 攻击创造条件，其中很多都与上文提到的漏洞有关。例如，可以利用数据库平台漏洞来制造拒绝服务攻击，从而使服务器崩溃。其他常见的拒绝服务攻击技巧包括数据破坏、网络泛洪和服务器资源过载（内存、CPU 等）。

8. 数据库通信协议漏洞

数据库通信协议是数据库的客户端和服务器端的通信所遵循的规则。在所有的数据库通信协议中，发现了越来越多的安全漏洞。在两个最新的 IBM DB2 Fix Pack 中，七个安全修复程序中有四个是针对协议漏洞的。同样地，最新的 Oracle 季度补丁程序所修复的 23 个数据库漏洞中有 11 个与协议有关。针对这些漏洞的欺骗性活动包括未经授权的数据访问、数据破坏以及拒绝服务。例如，SQL Slammer 2 蠕虫就是利用了 Microsoft SQL Server 协议中的漏洞实施拒绝服务攻击。

9. 薄弱的身份验证方案

薄弱的身份验证方案可以使攻击者窃取或以其他方法获得登录凭据，从而获取合法数据库用户的身份。攻击者可以采取多种策略来获取凭据。

(1) 暴力攻击。攻击者不断地输入用户名/密码组合，直到找到可以登录的一组。暴力过程可能是靠猜测，也可能是靠复杂的算法来破解用户名/密码组合。通常，攻击者会使用自动化程序来加快暴力攻击的速度。

(2) 社会工程攻击。攻击者利用人天生容易相信别人的倾向来获取他人的信任，从而获得其登录凭据。例如，攻击者可能在电话中伪装成一名 IT 技术经理，以“系统维护”为由要求提供登录凭据。

(3) 直接窃取凭据。攻击者可能通过抄写即时贴上的内容或复制密码文件来窃取登录凭据，例如 ATM 机中的窃取个人密码的装置等。

10. 备份数据泄露

若没有对备份数据库存储介质采取一定的安全措施，对于攻击者是毫无防护的。在若干起著名的安全破坏活动中，都是数据库备份磁带和硬盘被盗。为防止备份数据暴露，所有数据库备份都应加密。实际上，某些供应商已经建议在未来的 DBMS 产品中不支持创建未加密的备份。建议经常对联机的生产数据库信息进行加密，但是由于性能问题和密钥管理问题，这一方法通常是不现实的。

解决上述安全威胁的方法主要有：数据库安全访问策略、数据库水印技术、数据库安全管理，这些内容将在下面章节中详细介绍。

10.3 数据库安全访问策略

数据库得以安全访问的重要策略在于对各种访问加以控制，从而达到防止非法用户进入系统及合法用户对系统资源的非法使用的目的。数据库的安全控制技术主要有信息流向控制、推导控制、访问控制，其中访问控制技术的应用最广泛且最有效。

10.3.1 访问控制技术

访问控制(access control)是通过某种途径显式地准许或限制用户的访问能力及范围，以限制对关键资源的访问，防止非法用户的侵入或者合法用户的不慎操作所造成的破坏。访问控制技术保证了用户在对数据库操作之前必须先经过授权，这是数据保护的前沿屏障。数据库安全访问技术的研究内容主要有：自主访问控制、强制访问控制、基于内容的访问控制、基于精细粒度的访问控制、基于角色的访问控制以及使用控制等。除使用控制外，其他访问控制技术在当前的主流商用数据库系统中都已经得到了应用。

访问控制系统一般包括三个角色，即主体、客体和安全访问策略。其中主体是发出访问控制、存取要求的主动方，可以是用户或应用程序的进程；客体是被调用的程序或欲

存取的数据;安全访问策略是一套数据库访问的规则。

1. 自主访问控制

自主访问控制(Discretionary Access Control,DAC)是指系统根据主体是否具有对客体的所有权或衍生的访问权来决定主体是否能访问客体。当主体具有某种访问权、同时又拥有将该访问权授予其他用户的权利时,能够自行决定将其访问权直接或间接地转授给其他主体。在自主访问控制中,系统用户对数据信息的存取控制主要是基于对用户身份的鉴别和存取访问规则的确定。当用户要执行某项操作时,系统就根据用户的请求与系统的授权存取矩阵进行匹配比较,通过则允许该用户的请求,对其提供可靠的数据存取方式,否则拒绝该用户的任何请求。DAC 的管理简单灵活,但在安全性上存在漏洞。如某个获得访问权的主体可在客体的所有者不允许的情况下,将对客体的访问权转授给其他主体。

2. 强制访问控制

强制访问控制(Mandatory Access Control,MAC)是指系统根据主体被信任的程度和客体所包含信息的机密性来决定主体对客体的访问权。在强制访问控制下,数据库系统给所有主体和客体分配了不同级别的安全属性。无论数据如何复制,数据和其安全级别是一个不可分的整体,只有符合安全级别要求的用户才可以操纵数据。它禁止了拥有高安全级别的主体更新低安全级别的数据对象,从而防止了敏感数据的泄露,提高了数据的安全性。而且,用户不能以任何方式修改自身或任何客体的安全属性,因此就无权将任何资源的访问权赋予别的用户,只有特定的系统权限管理员才能根据系统实际的需要修改系统的授权状态,从而消除了 DAC 中的安全漏洞。

3. 基于内容的访问控制

基于内容的访问控制要求存取控制取决于数据的内容。目前,常被采用的基于内容访问控制机制是视图机制。进行存取权限控制时,可以为不同的用户定义不同的视图,把数据对象限制在一定的范围内,使机密数据不出现在不应看到这些数据的用户视图上。这样通过视图机制可以把要保密的数据对无权存取的用户隐藏起来,从而对数据提供一定程度的安全保护。

4. 基于精细粒度的访问控制

基于精细粒度的访问控制来源于高安全级别的多级关系数据库系统的设计需求。此种数据库中存放着安全级别不同的数据,其中的安全级别标记粒度可以是关系、记录或属性,但这些定义方式或者会使某些敏感数据的安全级降低,或者会使某些非敏感数据的安全级升高。因此,基于精细粒度的访问控制被提出。根据控制对象的粗细程度,访问控制可分为粗粒度和细粒度两种,通常把规定访问整个数据库表或由基本表导出的视图的某个层称为粗粒度的访问控制,而细粒度控制则是把安全控制细化到数据库的行级或列级。朴素的精细粒度权限解决方案是为每个需保护的元组定义一张视图;另外一

种解决方法是采用虚拟隐私数据库。

5. 基于角色的访问控制

基于角色的访问控制(Role-Based Access Control,RBAC)提供了解决具有大量用户、数据库客体和访问权限系统中的授权管理问题。RBAC 涉及用户、角色、访问权以及会话等主要概念,如图 10-2 所示。

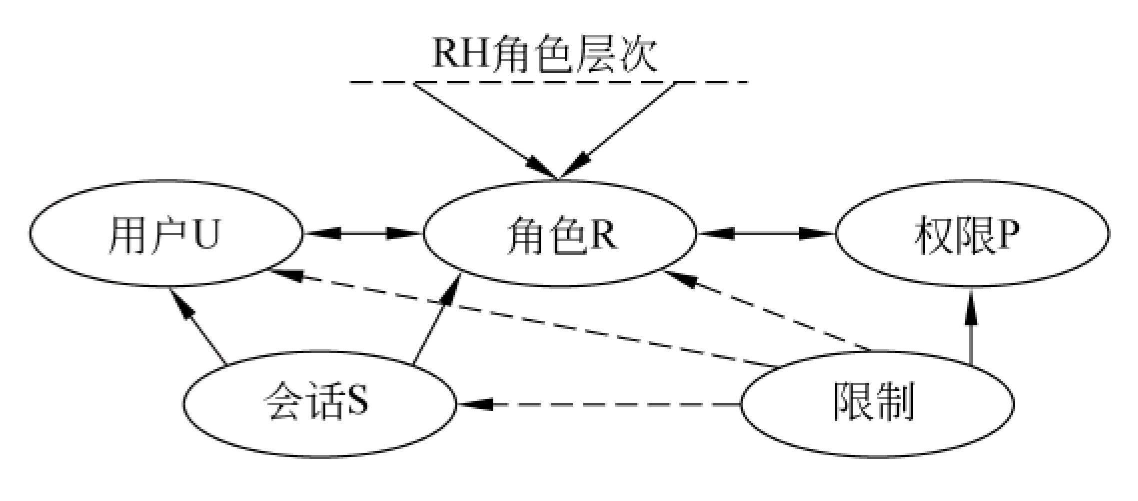


图 10-2 RBAC96 模型

角色是一组用户和一组操作权限的集合,角色中的用户可以执行这些操作权限。在数据库中创建一个新用户时,需要为其指定角色。用户与角色间、角色与访问许可权间都是多对多的关系。当用户登录到 RBAC 系统时,会有一个会话,此会话可能激活的角色是该用户全部角色的一个子集。角色可以根据实际需要生成或取消,用户也可以根据需要动态激活拥有的角色,这样就避免了用户无意间对系统安全造成的危害。由于数据库应用层角色的逻辑意义更为明显和直接,因此,RBAC 非常适用于数据库应用层安全模型。

6. 基于使用的访问控制

在传统的访问控制模型 DAC、MAC 和 RBAC 中,授权发生在访问前,而在整个访问期间可能需要对相对长期的访问或者访问权限进行立即回收,为了解决上述问题,使用控制模型被提出。使用控制模型(Usage Control,UCON)由主体、主体属性、客体、客体属性、权限、授权、证书以及条件 8 个部分组成,如图 10-3 所示。使用控制模型通过增加主体属性、客体属性、证书和条件解决了传统访问模型中存在的问题。

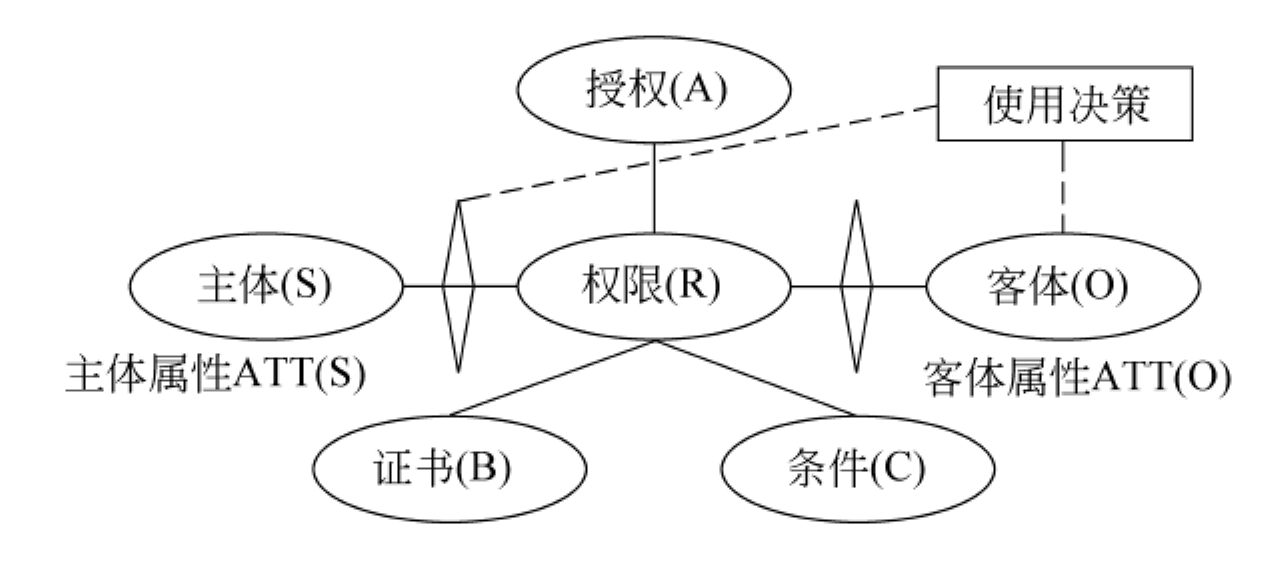


图 10-3 使用控制模型授权过程

10.3.2 数据库其他安全访问策略

在一般计算机系统中,安全措施是一级一级层层设置的。当用户要求进入计算机系

统时，系统首先要对用户的身份进行核实，就用到了用户标识和鉴别技术。除了系统保护外，还有防火墙等防止网络非法访问的安全技术。这里只讨论与数据库相关的用户标识和鉴别技术。

用户标识和鉴别是系统提供的最外层安全保护措施。其方法是由系统提供一定的方式让用户标识自己的名字或身份。每次用户要求进入系统时，由系统进行核对，通过鉴定后才提供使用权。对于获得使用权的用户若要使用数据库时，数据库管理系统还要进行用户标识和鉴定。用户标识和鉴定的方法有多种，而且在一个系统中往往是多种方法并举，以获得更强的安全性。

(1) 用户标识：用一个用户名或者用户标识号来表明用户身份。系统内部记录着所有合法用户的标识，系统在用户请求服务时鉴别此用户是不是合法用户，若是，则可以进入下一步的核实；若不是，则不能使用系统。

(2) 口令：为了进一步核实用户，系统常常要求用户输入口令。为保密起见，用户在中断上输入的口令不显示在屏幕上，通常以“*”或“·”代替。系统核对口令以鉴别用户身份。

10.4 数据库水印技术

10.4.1 数据库水印分类

数据库水印是指用信号处理的方法在数据库中嵌入不易察觉且难以去除的标记，在不破坏数据库内容和可用性的前提下，达到保护数据库安全的目的。

数据库水印主要可以分为两类：鲁棒性水印和脆弱性水印。鲁棒性水印主要应用于数据库的版权保护，攻击者可能会试图清除水印信息或者使水印信息不可检测，但保持嵌入水印的数据可用，因此这种水印要求具有非常强的鲁棒性，能抵御各种恶意攻击；而脆弱性水印则主要应用于数据的篡改检测，攻击者可能会试图修改嵌入水印后的数据而保持嵌入的水印信息不被改动，脆弱性水印要求对数据的修改非常敏感。

10.4.2 数据库水印的技术要求

理想的数据库水印技术应该充分考虑到关系数据库自身的特殊性以及各种攻击方式。其技术要求如下：

(1) 可嵌入要求：由于数据库中的冗余非常小，因此在关系数据库中难以找到合适的水印嵌入位置，只能在不影响原始数据可用性的基础上嵌入水印信息，而且由于关系数据库中的数据一般具有很强的语法结构和语法意义，在水印嵌入时不得与原有的语法结构和语法意义相矛盾，即有可能某些数据是不能修改的，这样的数据也就不能嵌入水印。

(2) 可操作要求：关系数据库中的数据通常需要进行一系列的运算，那么在进行一系列的运算后，水印仍然要附着于关系数据库的数据中，这是数据库水印研究的难点之

一。要求水印数据完全融入数据库中的数据中，均匀分布于整个数据库。

(3) 动态性要求：数据拥有者对带有水印的数据库进行更新时，水印信息应该随着数据的更新而嵌入，且更新数据的水印信息应与原数据库中的水印信息保持一致。当有新的数据加入时，水印信息应能实时嵌入；有数据修改时，水印信息不丢失；数据正常删除时，要保证数据库中水印信息的完整性。

(4) 盲检性：水印检测时，既不需要原始的水印信息，也不需要原始数据就可以从关系数据库的数据中检测提取出水印信息，实现水印的盲提取。这样确保非法复制的数据库副本中的水印总能被检测出来，而不需要依靠可能已经更新的原始数据库，这些副本可能被进一步的数据整合，与原始数据库可能有比较大的不同。

(5) 可管理性：带水印数据形成的数字产品，转移到带有水印管理功能或者兼容的数据库管理系统中仍然能够运行，水印可以随之迁移，不会轻易丢失。

(6) 二次水印问题：版权人对原始数据嵌入水印信息以后，数据产品可发布与其他人共享。这时，如果非版权人拿到数据后再对数据添加水印，那么，原始数据中既有版权人的水印信息，也新添上了非版权人的信息，如果两次的水印信息都有效地保留于原始数据上，一旦发生版权纠纷，就很难辨明谁才是真正的版权人了。还有更为更糟糕情况，就是如果非版权人使用了与版权人相同的水印嵌入算法加上自己的水印后就可能将版权人的水印信息替换掉了。

还有另外一种可能，非版权人使用了版权人数据的一部分，之后他在此基础上增添了自己的数据，水印技术需要对这两部分数据分别进行保护，水印信息之间的互不干扰成为一大难题，这对水印的嵌入、检测算法提出了更高的要求。

10.4.3 数据库水印的攻击

数据库在正常的维护更新中，常常需要删除数据、插入数据或更新数据。因而，对数据库水印的鲁棒性要求较多媒体数字水印更高。嵌入在数据库中的水印信息不能因为对数据库的常规操作而丢失，否则，水印技术在数据库的版权保护变得毫无意义。除此以外，数据库水印还应防御各种各样的恶意攻击。常见的数据库水印恶意攻击有：

(1) 子集修改攻击：攻击者希望通过修改数据库中的部分元组，去除数据库中的水印信息。

(2) 子集选取攻击：攻击者不使用数据库中的全部属性和元组，希望通过删除数据库中的部分元组，去除水印信息。

(3) 子集增加攻击：攻击者通过向数据库中添加元组的方式去除水印信息。

(4) 混合和匹配攻击：攻击者从多个类似的数据库中选取元组，创建自己的数据库。

(5) 可逆性攻击：攻击者在窃取的数据库中发现了一个随机出现的虚幻水印，就声称该数据库归他所有。

(6) 添加攻击：攻击者在已经加有水印的数据库中再嵌入自己的水印信息，并声称自己对该数据库具有所有权。

10.4.4 数据库水印算法

近几年来，国内外研究者在数据库水印方面取得了一些新的进展。较著名的有 IBM Almaden 研究中心的 R. Agrawal 和 J. Kiernan 对数据库进行的水印的嵌入和攻击试验。它针对一个特定的数据库，其中只包含数值型数据，且假定每个字段都能够添加水印，然后依据水印密钥和关键字确定需标记的字段及位置。还有美国 Purdue 大学的 R. Sion 等提出的对关系数据集合和数据库添加水印的方法，是基于"均方差"特性对数值型字段进行标记的。

目前常用的数据库水印算法主要包括以下两种。

1. 利用一定失真范围内的数据变形来嵌入水印

P. Agrawal 等于 2002 年首次对关系数据库进行了嵌入比特位模式的实验。该实验利用数据库关系中数值型元组存在的冗余空间，通过在某些数值型属性值中引入少量的误差，对其最低有效位(Least Significant Bits，LSB)进行位操作，实现水印信息的嵌入。

其基本思想是首先假设数据库中的数值型属性可接受一定程度的误差，只要改变在误差范围内，就不会影响数据库的正常使用。水印嵌入时根据用户给定的密钥和元组主键值以及可以标记的元组比例来确定哪些元组需要标记，并根据可以标记的属性和比特位数确定标记的属性及其比特位位置。然后将关系数据库中符合条件的某些元组的某些数值型属性值的比特位值按规则置为 1 或 0，作为一个标记。这样，在整个数据库中许多个比特位标记组合的比特位模式就是嵌入的水印信息。提取时先做相同的工作，确定标记的位置，再记录符合规则的元组总数，然后与由置信因子 Q 决定的阈值 T 比较来判断数据库的版权。

其中需要加标记的元组、元组的属性、属性的比特位位置以及具体的比特值都是由密钥、元组主键值和需要标记的元组比例控制算法来决定，这里密钥、元组标记比例、可标记属性数和比特位数只有关系数据库的所有者才知道。

2. 基于元组排序和划分集合实现水印嵌入

该方法由美国普渡大学的 R. Sion 等提出。首先根据元组的加密键值哈希对其进行秘密排序，然后基于"均方差"特性构造子集，取连续序列数据作为嵌入水印的基本单位。通过调整关键属性数据改变连续序列数据的分布特征来表示 1 和 0。R. Sion 等基于该方法开发了一个名为 WMDB 的数据库水印程序包，显示了较好的透明性和抗攻击能力。

其基本思想是对数值型属性进行标记的。给定数值型项目集合 $S=\{s_i,\cdots,s_n\}\subset R$，和一个秘密的排序密钥 k_s，首先根据标准化项目的最重要比特位的加密键值 Hash 对其进行秘密排序，例如，$\text{index}(s_i)=H(k_s,\text{msb}(\text{NORM}(s_i)),k_s)$。然后构造子集 S_i(即实现分组)用来嵌入比特位水印标记。假定水印信息有 m 个比特位长，则整个水印带宽将是 m 个比特位，每个比特位嵌入/隐藏到每个标记的 S_i 中。检测时需要用到嵌入时记录的子集信息。

上述两种数据库水印算法各有其优缺点。第一类方法采用基本的 LSB 嵌入算法，易

于实现,但水印信号的抗攻击能力较弱,而且难以嵌入有实际意义的水印信息。第二类方法具有较好的鲁棒性,但如果数据库中不同字段的取值范围相差较大,将导致计算获得的值只能对部分数据项适用,限制了水印嵌入的容量。

10.5 数据库安全管理

10.5.1 数据库安全管理要求

数据库的安全性很大程度上依赖于数据库管理系统。大多数的数据库管理系统是以操作系统文件作为建库的基础,所以操作系统安全特别是文件系统的安全便成为数据库管理系统安全的基本要求。因此,对数据库的安全管理可从数据库管理系统的安全运行管理和数据库管理系统中存储、传输和处理数据信息的管理着手展开研究。针对这两种安全管理,主要有加密和审计两种技术。

10.5.2 数据库加密技术

数据库加密技术的基本思想跟数据加密是一致的。它根据一定的算法将原始数据变为不可直接识别的格式,从而使得不知道解密算法的人无法获知数据的内容,达到数据库数据信息的安全管理。主要的加密方法有系统中加密、DBMS 内核层(服器端)加密和 DBMS 外层(客户端)加密三种。

1. 系统中加密

将数据先在内存中进行加密,然后文件系统把每次加密后的内存数据写入数据库文件中去,读取时再逆操作进行解密。

1)数据文件存储加密

文件型数据库系统是以文件的形式进行存储的,因而可以使用加密文件中数据的方法来加密数据库。首先将存放在内存里的数据使用合适的加密算法进行加密,然后对加密后的内存数据以数据库文件的形式存储在外部存储器中。需要使用数据时,只需要对数据库文件中的数据进行解密即可。这种加密的方法相对比较简单,只要妥善保管密钥,就能够保证数据库的安全。但是,每次读写数据库都要进行加密或解密,相对比较麻烦,并且会影响数据库操作执行的效率。

2)数据库对象加密

数据表中进行数据存储的最小单位是数据项。因此,可以考虑对数据项加密来获得高安全性。采用数据项级存储加密的方法将数据库中不同的记录、每条记录的不同字段都采用不同的密钥加密。但此方法同样大大降低了数据库存取的效率。而数据库数据的选择性加密方法,只对敏感信息进行加密,可以有效地避免频繁加解密操作对数据读取速度的影响,从而有利于用户在效率与安全性之间达到平衡。

2. DBMS 内核层加密

在 DBMS 内核层加密时需要对数据库管理系统本身进行操作。这种加密是指数据在物理存取之前完成加解密工作。这种加密方式的优点是加密功能强,并且加密功能几乎不会影响 DBMS 的功能,可以实现加密功能与数据库管理系统之间的无缝耦合。其缺点是加密运算在服务器端进行,加重了服务器的负载,而且 DBMS 和加密器之间的接口需要 DBMS 开发商的支持。

3. DBMS 外层加密

DBMS 外层加密将数据库加密系统做成 DBMS 的一个外层工具,根据加密要求自动完成数据库数据的加解密处理,加解密运算可在客户端进行,它的优点是不会加重数据库服务器的负载,并且可以实现网上传输的加密,缺点是加密功能会受到一些限制,与数据库管理系统之间的耦合性稍差。

10.5.3 数据库审计技术

审计功能把用户对数据库的所有操作自动记录下来放入审计日志中,以备系统管理员分析系统的访问情况,以及违反规则之后做追查责任之用,达到数据库运行的安全管理。

审计记录包括以下信息:事件发生的日期和时间、用户、事件类型、事件是否成功。当系统检测到有危害系统安全的事件发生时,可以设置系统发出自动报警信息,同时执行一系列的操作,阻止该用户的非法操作。由此,可以有效防止来自外部的对用户计算机文件的恶意窃取。另外,系统管理员可以利用审计跟踪的信息,重现导致数据库现有状况的一系列事件,找出非法存取数据的人、时间和内容等。

审计一般可以分为用户级审计和系统级审计。用户级审计是任何用户可设置的审计,主要是用户针对自己创建的数据库表或视图进行审计,记录所有用户对这些表或视图的一切成功和(或)不成功的访问要求以及各种类型的操作。系统审计职能由系统管理员设置,用以监测登录要求的成功或失败,以及其他数据库级权限下的操作。

但审计通常是很费时间和空间的,所以 DBMS 往往都将其作为可选特征,允许系统管理员根据应用对安全性的要求,灵活地打开或关闭审计功能。审计功能一般主要用于安全性要求较高的部门。

思　考　题

10.1　什么是数据库的安全性?

10.2　如何对用户账号进行授权管理,一般有哪些权限?请简要说明。

10.3　为什么要进行数据备份?数据库备份包括哪些主要内容?

10.4　什么是强制访问控制机制?强制访问机制如何防止“特洛伊木马”的非法访问?

10.5 什么是自主访问机制？自主访问控制机制与强制访问控制机制的区别有哪些？

10.6 数据库水印主要类型有哪些？

10.7 数据库水印与多媒体数字水印的区别是什么？什么是盲检测？

10.8 列出几种典型的数据库水印算法。

10.9 数据库加密技术有哪几种方法？

10.10 数据库审计技术的优缺点是什么？

10.11 即使将纵向奇偶位作为错误校验码，但仍然检测不到对数据库的篡改，为什么？(纵向奇偶位计算每个字节的第 n 位，一个奇偶校验位对所有第 0 位计算并保留其值，另一个奇偶校验位对所有第 1 位计算，等等。)

10.12 多级安全数据库管理系统使用加密技术的目的是什么？

10.13 如何通过可信任操作系统提供给用户的多级分离来实现数据库管理系统？

10.14 假定操作系统已经有安全级别 r，r 是 C2 或 B1 或 B3 等。基于对操作系统的信任，定义一个数据库管理系统的信任度策略，并加以解释。

参考文献

[1] 粟松涛，李春文，孙正顺. 一种新的 B/S 系统权限控制方法. 计算机工程与应用. 2002，38(1)：99-101.

[2] 崔艳荣，文汉云. 数据库安全模型及其应用研究. 计算机应用研究，2005，22(7)：146-147.

[3] Park J，Sandhu R. The UCONABC usage control model. ACM Transactions On Information and system Security，2004，7(1)：l-47.

[4] 严和平，汪卫，施伯乐. 安全数据库的推理控制. 软件学报，2006，17(4)：750-758.

[5] 朱虹，汪皓，薛慧，等. 异构平台的数据库安全技术. http://www. eew. eom. en/cidresea rchhnfo/htm2004/2004. 1 119-10JEW. asp.

[6] 黄敏，张浩，曹加恒. 一种基于关系数据库的水印技术. 计算机工程与应用，2005，25(10)：153-199.

[7] 张勇. 水印关系数据库研究. 中国人民解放军理工大学博士论文，2004. 5.

[8] 武荣，曹加恒，黄敏，等. 关系数据库的数字水印新技术. 武汉大学学报(理科版)，2005，51(5)：589-593.

[9] 周旭，毕笃彦. 基于中国剩余定理的 GIS 数字水印算法. 中国图像图形学报，2004，9(5)：611-615.

[10] R. Agrawal，J. Kiernan. Watermarking Relational Databases. In：Proceeding of the 28th VLDB Conference，Hong Kong，China，August 2002.

[11] Sion R，Atallab M，Prabhakar S. Rights protection for relational data. IEEE Transactions on Knowledge and Data Engineering，2004，16(12)：98-109.

[12] WMDB System Architecture. http://www. cs. stonybrook. edu/～sion/projects/wmdb，2004.

[13] 张子谦. 数据库隐通道安全防护策略研究. 南京航空航天大学硕士学位论文. 2008.

[14] 向阳，魏玉鹏，王改梅. 数据库访问控制技术研究. 河南科技学院学报(自然科学版)，2006，34(2)：101-104.

[15] 熊燕妮，黄铂. 浅谈数据库访问控制技术. 武汉生物工程学院学报，2009，5(1)：27-29.

[16] 黄飞，黄正东，王琳. 当代信息技术条件下数据库安全技术研究. 医院数字化，2010，31(10)：51-54.

［17］ 王永，刘秀军，马建峰，访问控制模型分析. 晋中师范高等专科学校学报，2002，19(2)：109-112.
［18］ 韩言妮，刘国华，沈兵红. 数据库层上的细粒度访问控制技术. 燕山大学学报，2006，30(4)：345-348.
［19］ 朱勤，于守健，乐嘉锦. 数据水印研究与进展. 计算机工程与应用，2006，42(29)：198-201.
［20］ 信息安全技术——数据库管理系统安全技术要求，GBT 20273—2006.
［21］ 孔令美. 浅谈数据库安全管理. 电脑知识与技术，2009，5(11)：2804-2806.
［22］ 陈明刚. 基于数值属性的关系数据库水印研究. 湖南大学硕士学位论文，2007.